1. REPORT DATE **APR 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE		5a. CONTRACT NUMBER
Three Papers in International Health Policy. Modeling the Links Between Economics and Epidemiology		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rand Corporation,Pardee RAND Graduate School ,1776 Main Street,Santa Monica,CA,90407-2138**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**187**	

Three Papers in International Health Policy

Modeling the Links Between Economics and Epidemiology

Arindam Dutta

RAND PARDEE RAND GRADUATE SCHOOL

Summary

Economic and epidemiological models should be connected such that resource allocation decisions in international health policy are based on realistic expectations of disease outcomes, and meet public finance criteria. However, epidemiological models which best capture uncertain real-world processes have been in the specialist domain and are usually not amenable for use by the informed social scientist or policymaker. As a result certain international health policy decisions have had uncertain bases and related economic analyses have been wildly imprecise. Desktop models can be constructed which closely match the dynamics of complex epidemiological simulations but are easily editable with new data such as: proven results from clinical trials or other laboratory testing on the efficacy of certain interventions, and indicative bounds from mathematical modeling for outcomes with or without interventions in a real-world setting.

Paper I establishes the benefits of linking epidemiological modeling with international health resource allocation decisions, reviewing the recent modeling literature on pandemic influenza control. The review indicates that outbreaks in resource-poor settings are controllable with moderate resource intensity and complexity of effort for viral strains of moderate infectiousness. However, very high resource allocations for preparedness in industrialized nations – at low geographic risk for the pandemic – are predicated on containment failure in countries at higher risk of outbreaks. Without assuming the infectiousness of a future flu virus, a redistribution of resources to the developing countries at primary risk reduces overall systemic risk of containment failure. The payoffs in terms of reduced global mortality and morbidity are higher with increased infectiousness.

The two other papers are associated with implementing the experimental desktop models for the context of India. *Paper II* first constructs a scenario-based non-epidemiological model of pandemic influenza introduction to, and subsequent spread within India under various assumptions. The model uses published data on attack rates in Asia during previous pandemics as well as seasonal influenza. The model exploits geographical risk variations across provinces of India as well as the provinces' demographics, transport networks, and rural-urban settings. The simulated epidemics achieve overall attack rates of 7-21% of the population, well below the 30% figure used in

economic analyses. A range of benefits is estimated for preventing introduction of the virus via the air network (as compared to the later, inevitable entry via land & sea), from 60,000-100,000 averted deaths and averted economic costs of US$400-710 million. Separately, a desktop spatial epidemiological model – EpiFlex – is exploited to analyze disease spread in New Delhi. The attack rates under a realistic metapopulation simulation are higher than using the average-based attack rates in the countrywide model. The EpiFlex results also indicate that superspreading, which resists conventional outbreak containment, might be a factor in the dense urban environments of India. Policy recommendations include the inclusion of a proactive air travel shutdown policy, with potential social benefits higher than total private costs.

Paper III re-estimates the estimates of people living with HIV/AIDS (PLWHA) in India by combining the available prevalence data from the latest sero-surveillance data as well as the National Family Health Survey (NFHS-3) of 2005-06. The estimated total prevalence after accounting for biases in 2008 is 0.4%, within the bounds of the recent official announcement of 0.36% for year 2007 (0.27%-0.47%). The related 2.41 million PLWHA confirms that the prevalence had been overestimated by 230% till 2007.

The paper continues to comprehensively analyze antiretroviral (ARV) policy in India, beginning with the estimation of total costs of utilization under public and private market rates for first-line ART. A cohort simulation is conducted using a desktop model of disease progress in the population without access to ARVs. Deaths and morbidities are measured, and the availability of opportunistic infection prophylaxis is accounted for. Separately, the same modeling framework is used to simulate disease progress in a cohort of 122,947 (based on those eligible to ARVs using WHO criteria from the 15% of those without any treatment currently). First-line combination ARV therapy based on non-nucleoside reverse transcriptase inhibitors (NNRTI) is administered to this cohort with different treatment variations, and treatment efficacy is simulated based on published clinical and observational data. Second-line protease inhibitor (PI) therapy is made available, and outcomes are revised.

Cost-effectiveness (CE) is estimated based on years of life saved compared to no treatment. The desktop model results confirm cost-effectiveness of both one and two lines ART compared to no treatment, and CE values are within bounds of a recent stochastic, individual-based simulation (Freedberg *et al.* 2007), as well as prior studies. Advantages of

the model – besides tractability and customizability – includes the inclusion of treatment failure and the ability to model physician decision to continue first-line treatment even on failure, a reality considered in India given the low availability of second-line treatment. Here, the model yields novel results: continuing first-line therapy till end of follow-up (the realistic treatment horizon) is cost-effective compared to ending first-line treatment at physician-determined treatment failure. Compared to Freedberg *et al.*, PI-based second-line ART is not found cost-effective by WHO standards compared to NNRTI-based first-line therapy, at prevailing PI-combination drug costs.

Policy recommendations include expanding first-line ART coverage, which is cost-effective, and will avert many premature deaths and reduce severe morbidity. Increasing survival with second-line ART comes at higher cost. Co-trimoxazole prophylaxis in situations without ART is an interim necessity. Continuing, compared to stopping, first-line ART in the case of treatment failure is promising, as long viral resistance and transmission is monitored and curtailed.

Acknowledgments

Table of Contents

Foreword to the Three Papers

The interaction between microorganisms and humans is an ancient one, which had only a few definable outcomes for the individual till very recently in our history: illness, recovery, or death. The response of communities to infectious disease varied across cultures and levels of scientific knowledge, but was also limited, even as it introduced its own uncertainty in the levels of those few definable outcomes. With the use of a complex suite of modern prophylaxis, treatment, and care methods, and the ever-increasing complexity and intensity of the public health response to infection risk, the levels of uncertainty increased, though on average recovery is more common for the individual and the well-understood disease. An individual's outcome for such a disease now depends on what public and private health environment she is exposed to. If it is important to predict those outcomes, then it is required that we capture the effects of the health environment. Clinical trials tell us what to expect from pharmaceuticals, even given the caveat that the individual experience may vary. Observational studies track the disease course in groups of individuals on treatment, and this informs clinical management for yet other groups. Public health measures are studied historically, as well as prospectively, via simulations. There is an attempt to divine the future whenever man meets virus, but it is a cloudy crystal sphere.

It is uncertain if successes such as the one against smallpox can be repeated; efforts are still underway to complete the victory against polio more than fifty years after the vaccine was developed. As a separate source of uncertainty in outcomes, harmful microorganisms have responded to our pharmaceutically-mediated desire for their extinction. At a species level, the microorganism's engine of evolution operates across the interaction of host, virus/bacteria, and the pharmaceutical involved. As a result, resistance develops in time, and even well-understood treatments do fail. Also, new diseases emerge – even if based on an old pattern, such as the plague – that are difficult for modern methods to counter. Uncertainty again gets a boost.

Neither infectious disease nor uncertainty over its outcomes will go away. However, it is important that such outcomes be predicted, since they are the basic input into decision-making over resources that must be allocated in the perpetual war against the microorganism. The pressure on resources is intense; while the list of microorganisms defeated and relegated to history is distressingly short, the list of public health demands is not. In the developing world, the number of individuals who desire a longer and healthier

life has grown with increasing prosperity and knowledge; and the conditions under which governments are obliged to satisfy this desire have also grown: democracy, better governance, improving primary health systems, and a 'global' mindset in viewing disease risk. Global development assistance for health (DAH) increased from US$2.5 billion in 1990 to US$14 billion in 2005, about 13% of all official development assistance[1]. While some emergent risks –not yet fully realized – have seen a surfeit of new spending (more than US$1 billion in donor funds for pandemic influenza preparedness since 2004), diseases once thought controlled have re-emerged as a risk, but have been relatively neglected (e.g., river-blindness, which has shown signs of resistance to its mainline treatment[2]). Elsewhere, HIV/AIDS continues to be important in the new scale-up of DAH. Its share in DAH went from 8% to 21% over 2000-04, while primary care, the first line of defense against childhood and some adult diseases, saw its share fall from 28% to 15% over 2000-04[3].

The need for public preparedness spending on emergent risks such as pandemic influenza is taken seriously based on studies that suggest an increasingly interlinked, skills-based world economy cannot afford to lose a large number of people and a shutdown of economic activity during a pandemic. Other studies have pointed out that the avoidance behavior of individuals when an outbreak of a highly communicable and virulent disease such as pandemic influenza starts may itself be the cause of major loss to demand[4].

Separately, the case for continued public health spending on HIV/AIDS treatment (as compared to prevention) is made based on the consequences of inattention. First, a sexually-active adult with HIV+ is still in a position to spread the disease to at least one or two others depending on the individual and the social context, thus imposing avoidable negative externalities. Treatment may cause some individuals to take more risks – though this is culturally conditioned – but it reduces the viral load in all patients who respond to effective treatment, and also ensures the patient remains in behavioral counseling. Second, the treatment of the disease (though not a cure) imposes long-term economic burden on those who must pay, if not for the drugs, then for enhanced nutrition and caregivers' income. This is difficult for low-income individuals. Therefore, lack of public health support

---

[1] "Healthy development: the World Bank strategy for HNP results", April 24th, 2007, Washington DC: The World Bank
[2] "River blindness resistance fears," 14th June 2007, BBC World, accessed March 21st 2008 http://news.bbc.co.uk/2/hi/health/6753003.stm
[3] The World Bank, op. cit.
[4] W. McKibbin & A. Sidorenko, "Global macroeconomic consequences of pandemic influenza," 2006, Sydney: Lowy Institute for International Policy

for treatment has distributional consequences. Finally, the lengthy symptomatic disease period implies excess suffering. Individuals with HIV, when asked have expressed a high willingness-to-pay for the treatment[5]. However, few can still afford treatment and associated costs out-of-pocket for the full disease course even despite recent decreases in the cost of antiretrovirals. If resistance to treatment is considered, the prospective costs rise: second-line drugs are still very expensive, even in the public sector.

The resource allocation question is important in most developing countries, but even more so for India, which carries a heavy burden of infectious disease-related mortality and premature death, even as other Asian countries have made rapid progress in achieving an epidemiological transition (a decline in infectious disease). South Asian governments in general spend below the average of all low-income countries on health as a percentage of total health expenditure (18.8% compared to 24%), while out-of-pocket payments are among the highest in the world (76% of total health expenditure)[6].

Though there are aspects to both pandemic influenza and HIV/AIDS that imply, or are documented to suggest, very high social and personal costs, the basis for continued public health resource allocations for these diseases should not be based on arguments of 'exceptional disease'. Many other infectious and even chronic diseases are debilitating and have a wide enough prevalence to cause macroeconomically significant losses to productivity and production: e.g., diabetes. Funding for health systems rather than particular diseases could also be an alternate priority. In this policy climate, for pandemic influenza, it is important to know what the plausible scale of future human losses might be which will then determine economic disruption and loss to output, which in turn will suggest the correct levels of preparedness spending. For HIV/AIDS, it is important that an objective criterion for the value of the treatment given a public health perspective is known and an estimate is found for the feasible size of any proposed scale-up of such spending. One of these objective criteria is cost-effectiveness as measured in terms of incremental spending per year of life saved (YLS), potentially adjusted for disability.

The need to answer these questions and to tie them to what are, in effect, issues of economic necessity, again brings us back to uncertainty and models. Pandemic influenza could affect millions of individuals and justify large outlays – but can this be known with

[5] I. Gupta, "Willingness to pay for antiretroviral therapy for HIV positive individuals in India," 2007, *Forum for Health Economics & Policy* 10:Economics of the HIV Epidemic, Article 7
[6] The World Bank, op. cit.

any surety beforehand? Could the best policy alternatives to prevent the spread of such a highly infectious disease be tested and assigned costs and benefits? This dissertation answers both of these questions, beginning with the results of recent modeling of pandemic influenza control interventions which have had a global perspective. These models attempt to capture the uncertainty over outcomes that different contexts of pandemic influenza outbreak, spread, and policy environments bring.

However, capturing all the elements that impose uncertainty over the outcomes of interest – deaths and illnesses with/without intervention – is also tantamount to creating pandemic influenza models of some complexity as well as costliness in effort and resources. Similarly, the simulation of the outcomes in a cohort of HIV+ individuals under antiretroviral treatment requires the mimicking of the risk of opportunistic infection, chronic mortality, and the inclusion of various cost heads, such as CD4 T-cell count testing. The likelihood that each individual's disease course will be different is another level of uncertainty that is captured if the models build for such level of detail. Such models are the current gold-standard in simulating the disease problem and prospecting for policy answers, but are very expensive to construct and somewhat opaque to the social scientist who would like to understand and apply them to various resource allocation contexts.

This dissertation attempts to 'simulate the simulations' using desktop models that capture the intent and the dynamics of the more complex/expensive and less tractable 'gold standard' simulations for both pandemic influenza and HIV/AIDS. In general, this is possible because of the path opened by the creativity and complexity of the gold standard modeling studies – which are discussed in detail in the following papers – but also because the desktop modeling method can be supported by the same clinical and laboratory data. The sacrifice is that some of the elements of uncertainty at the individual level can no longer be modeled. Instead the desktop models involve groups of individuals where the variation in disease-related characteristics of the group (average of all the individuals included in a group) is expected to yield an acceptable capture of some of the uncertainty in outcomes. The models are designed to be inherently customizable – when not based on customizable software in itself – and are intended to give policymakers another tool to estimate the cost-effectiveness of treatment for HIV, and the projected cost and benefit of pandemic influenza interventions in a highly uncertain world.

Paper I: The Effectiveness of Policies to Control a Human Influenza Pandemic: Review and Analysis

Abstract: The studies reviewed in this paper indicate that with adequate preparedness planning and execution it is possible to contain pandemic influenza outbreaks where they occur, for viral strains of moderate infectiousness. For viral strains of higher infectiousness, containment may be difficult, but it may be possible to mitigate the effects of the spread of pandemic influenza within a country and/or internationally with a combination of policies suited to the origins and nature of the initial outbreak. These results indicate the likelihood of containment success in 'frontline risk' countries, given specific resource availability and level of infectiousness; as well as mitigation success in 'secondary' risk countries, given the assumption of inevitable international transmission through air travel networks. However, from the analysis of the modeling results on interventions in the U.S. and U.K. after a global pandemic starts, there is a basis for arguing that the emphasis in the secondary risk countries could shift from mitigation towards containment. This follows since a mitigation-focused strategy in such developed countries presupposes that initial outbreak containment in these countries will necessarily fail. This is paradoxical if containment success at similar infectiousness of the virus is likely in developing countries with lower public health resources, based on results using similar modeling methodologies. Such a shift in emphasis could have major implications for global risk management for diseases of international concern such as pandemic influenza or a SARS-like disease.

Acronyms & Abbreviations

ACIP: Advisory Committee on Immunization Policy
AVE_I: Antiviral Efficacy in Reducing Infectiousness
AVE_S: Antiviral Efficacy in Reducing Susceptibility
BMR: Blanket Movement Restrictions
CDC: Centers for Disease Control (US)
CFR: Case Fatality Rate
DHHS: Department of Health and Human Services
GAR: Gross Attack Rate
GTAP: Geographically Targeted Antiviral Prophylaxis
H5N1: Hemagglutinin-5, Neuraminidase-1
H2N1: Hemagglutinin-2, Neuraminidase-1
HCW: Healthcare Workers
IATR: International Air Travel Restrictions
MIDAS: Models of Infectious Disease Agents Study
NPI: Nonpharmaceutical Intervention
NVAC: National Vaccine Advisory Committee
R_0: Basic Reproductive Number
RMR: Reactive Movement Restrictions
ROW: Rest of the World
SARS: Severe Acute Respiratory Syndrome
SE: Southeast
SEIR: Susceptible Exposed Infectious Removed
TLC: Targeted Layered Containment
TAP: Targeted Antiviral Prophylaxis
VE_I: Vaccine Efficacy in Reducing Infectiousness
VE_S: Vaccine Efficacy in Reducing Susceptibility
WB: The World Bank
WHO: World Health Organization

Glossary of epidemiological terms used

Asymptomatic transmission: For influenza, the transmission from person-to-person where the infecting individual does not manifest visible/detectable symptoms of influenza illness.

Basic Reproductive Number, R_0: The mean number of secondary cases caused by a typical single infectious case in a completely susceptible population (i.e., with no prior immunity), in the absence of public health interventions.

Case Fatality Rate*: The proportion of persons with a particular condition (cases) who die from that condition. The denominator is the number of incident cases; the numerator is the number of cause-specific deaths among those cases.

Chemoprophylaxis: Using antivirals or other drugs to prevent infection with the disease (can be pre- or post-exposure to the disease-causing agent, e.g., the influenza virus).

Epidemic: A generalized epidemic; or, a sequence of outbreaks in a large geographically defined area such as a city, a province, or a country.

Gross Attack Rate: The proportion, usually of a country population, that has a clinical case of the infectious disease during a defined period of time, such as the duration of the epidemic.

Herd Immunity*: The resistance of a group to invasion and spread of an infectious agent, based on the resistance to infection of a high proportion of individual members of the group. The resistance is a product of the number susceptible and the probability that those who are susceptible will come into contact with an infected person.

Incubation period (influenza): The longest period between the introduction of the virus into a host and the occurrence of the first clinical signs of the disease.

Outbreak: A localized epidemic; or, a sequence of related infectious disease cases in a geographically defined area such as a neighborhood, town, or city. Sometimes considered synonymous with epidemic.

Pandemic: An epidemic that is geographically widespread; occurring throughout a region or even throughout the world.

Physical prophylaxis: Using masks or other physical devices to prevention infection with the disease.

Prophylaxis: Prevention of infection with a disease, pre- or post-exposure to disease-causing agent.

Quarantine: The physical separation of healthy people who have been exposed to an infectious disease – for a period of time – from those who have not been exposed.

Social distancing: A disease prevention strategy in which public health authorities limit social (face-to-face) interaction to reduce exposure to and transmission of a disease. These limitations could include, but are not limited to, school and work closures, cancellation of public gatherings and closure or limited mass transportation.

Virulence*: The proportion of persons with clinical disease, who after becoming infected, become severely ill or die.

* Source: CDC Epidemiological Glossary. www.cdc.gov/reproductivehealth/EpiGlossary/glossary.htm

Introduction

The recent efforts for pandemic influenza preparedness at local, country, and international levels exemplify rapid, research-driven creation and execution of an agenda for a disease of international concern. A prominent part of the research on a potential pandemic of influenza has focused on strategies to prevent or mitigate such an event. Much of this research has been highly technical and not adequately surveyed such that policymakers can make choices fitting the context of their city or country. This paper reviews this field, focusing on policies to control – i.e., prevent or mitigate – a human pandemic of influenza, and presents the major conclusions from the literature.

A rationale for this literature review is that research on the effectiveness of control policies is driving multilateral and national strategies at three levels. First, the World Health Organization (WHO), and the EU and US agencies responsible for pandemic preparedness have recommended strategies (*sets of policies*) for controlling a human pandemic of influenza. The recommended actions in certain contexts aim at containing an outbreak rapidly before it can spread; in others the emphasis is on mitigating the overall attack rate. The WHO has recently issued an Interim Protocol for containment measures under the title *Rapid Operations to Contain the Initial Emergence of Pandemic Influenza*[1]. Recommended policies here substantially derive from insights from mathematical modeling studies[3,5] that compare policies for containing pandemic outbreaks. Second, there are studies focusing on particular policies, such as limiting international air travel, vaccination, or community-level policies. These studies have informed specific WHO guidelines, e.g., for recommending a restriction on flights from certain cities; as well as country governments' choices of implementing such guidance. Third, policymakers have learned from research on the 2003 SARS outbreak. The emphasis on these three background literature sources in this review follows the order in which they were mentioned.

Before the aims of this review can be discussed, it is necessary to clarify some nomenclature. In general, effectiveness for a containment–focused strategy refers to the reduction of the spread of pandemic influenza within a country or internationally. For a mitigation strategy, effectiveness refers to reducing the overall number of cases, i.e., the attack rate, and to delay and reduce the peak rate of cases per day. A mitigation strategy lowering the overall illness attack rate unambiguously reduces hospitalizations and deaths

that are related to human and economic losses. For two strategies achieving similar reductions, the one more feasible is more effective. In this review, effectiveness of a policy (used here interchangeably with 'intervention') is differentiated from baseline efficacy, as established in theory or in the laboratory. Efficacy is an important determinant of policy effectiveness in modeling studies. However, effectiveness accounts for likelihood of use, feasibility, transmission dynamics, etc., in an outbreak or an epidemic. Given the rationale above and the definition of terms, this review has three specific aims as specified below.

I. Link in one document the current research on the efficacy of specific policies; the properties of a potential pandemic-capable influenza virus relevant to its infectiousness and virulence; the assumptions and methods behind mathematical or other models of pandemic influenza spread and control; and the results of studies of the effectiveness, cost (or cost-effectiveness), and ethical dimensions of policies and strategies. For example, such linkage can help refer the claims of modeling studies to the baseline efficacy of particular policies, and to assess the generalizability of the results given modeling assumptions, the assumed properties of the virus, etc.

II. Provide a taxonomy of control policies and to collate as completely as possible the results of modeling studies that assess their effectiveness in specific or in combination. This allows the comparison of the effectiveness of the same policy when viewed in isolation, or as part of a strategy group of policies aimed at containment or mitigation. This difference is relevant if particular countries may afford only one or few of the policies rather than the most effective group of policies.

III. Further the debate on the appropriate international mix of funding for pandemic influenza preparedness that reduces the global risk of such an event. As per the current distribution of funding, countries with past outbreaks and at risk of future outbreaks are spending less per capita than some countries with no outbreaks but larger resources. This imbalance in preparedness spending is driven by large pre-investments in the latter group in mitigation policies such as vaccine development and antiviral stockpiling. These are resource intensive and hence disproportionately the preserve of wealthier countries. Such spending is driven by a logic summarized as 'hope for the best but prepare for the worst'. This assumes that containment policies after a pandemic outbreak in the frontline states (those with current or past outbreaks) will fail and hence the spread to the West is inevitable; and further that subsequent containment

policies at airports or other entry points in the West will also fail. Discussion of the modeling results below, especially from studies that address a global redistribution of resources[80], should illuminate this debate.

Ethical and cost considerations can make the choice complex between two similarly effective and feasible control policies. For example, some mitigation policies are selective in terms of who is pre-protected from infection, raising ethical issues that require review. Cost-effectiveness studies of policies for pandemic influenza have been rare, reflecting the difficulty in rigorously estimating the benefits or costs, or negotiating the ethical issues in valuing sickness and death. The few such studies available will be reviewed. A different angle is taken by studies that estimate the general benefit of preventive policies in terms of the avoided macroeconomic costs of a potential pandemic. Such studies of potential national and international losses to economic activity are beyond the scope of this review.

Immediately below, a section surveys what is currently known about the properties of a pandemic influenza virus that determine the ease of containment or mitigation in an outbreak. Also in the same section, the results from published reviews of efficacy for a few policies are summarized. The next section provides a broad overview of the major methods for modeling the transmission of pandemic influenza and for the prospective comparison of control policies. The main section of the paper follows, initially providing the taxonomy of interventions. On this basis, studies that compare various control policies – for containment and mitigation – are analyzed, and those that evaluate a specific policy. The penultimate section reviews cost and cost-effectiveness studies. Ethical issues are briefly discussed in the concluding section, which can be read as an executive summary of the review. The intended audience includes public health practitioners, policymakers, and researchers involved in health. No detailed knowledge of pandemic influenza epidemiology is assumed.

Determinants of a pandemic influenza outbreak and its control

The properties of a future pandemic influenza virus relevant to the containment and mitigation of outbreaks may be uncertain, but studies have attempted to create scenarios for policy analysis based on three sources of information. First, clinical, epidemiological, and laboratory data exist in the form of tissue samples from patients, records of public health measures, as well as overall morbidity and mortality estimates from influenza pandemics of

the 20th century: in 1918-20, 1957-58, and 1968-69. Second, researchers have been tracking the evolution of the current avian influenza A (subtype H5N1) virus and the associated human cases. This virus is considered the prime candidate for generating a pandemic capable strain. Predicting potential mutations in the H5N1 strain helps create new scenarios of the properties of a pandemic virus and inform prevention and control measures. Third, there are extensive studies of the transmission of seasonal influenza (various subtypes), which can help model the transmission of a more efficient human communicable strain of any avian influenza A virus that attacks people.

A discussion of how an avian virus of subtype H5N1 may potentially mutate to pandemic capable form is omitted. Assuming that this mutation occurs, a substantial proportion of the world population – some studies assume 60%[62] – would be susceptible. The two basic properties of a pandemic influenza virus relevant to prevention and mitigation are considered below: *infectivity* and *virulence*. Thereafter, some concepts related to infectivity – or more generally, transmissibility – used in modeling outbreaks are discussed with reference to pandemic influenza viruses (e.g., the effective contact rate, the basic reproduction number). Connected to both infectivity and virulence, the current science on antiviral, vaccine, and personal protective equipment efficacy is discussed.

Infectivity: The more intrinsically infectious a virus, the higher the likelihood that when an infected person meets an average uninfected person, a second infection will occur. Here, an infection is defined as developing antibodies to the viral presence in the body – also known as 'exposure' – and is prior to developing a clinical case of the disease (i.e., symptoms). Given a level of base infectivity of the virus, the likelihood of a secondary infection is modified by several factors - the effectiveness of the contact between infected and uninfected, environmental conditions, and the presence of preventive barriers (i.e., prophylaxis via pharmaceuticals, or physical protection). The issue of effective contact is briefly considered after *virulence* below.

There are three types of human influenza viruses: A, B, and C. Of these, only influenza A viruses, which have subtypes based on the hemagglutinin (HA) and neuraminidase (NA) protein combinations on the viral surface, have caused pandemics. Influenza B virus strains may be widely transmitted in seasonal epidemics, but usually do not cause pandemics. Influenza C virus generally causes mild illnesses in humans[2]. In this review, only influenza A is considered. The base infectivity of influenza A viruses depends

on the rate at which the average infected person 'sheds' the virus, the timing of the 'peak shedding' period, the mode of transmission for the shed virus, minimum size of an infectious dose of virus, and survivability of the virus on surfaces and in the air.

Table 1. *Determinants of the Base Infectivity of Influenza Virus Strains*

Characteristic	*Conclusions for seasonal influenza A*	*Implication for epidemic control*	*Review study source*
Virus shedding rate & duration	▪ High virus load in respiratory secretions ▪ Shedding up to 24-48 h *before* symptoms ▪ Virus shed for up to 5 days ▪ Volume & duration higher in children	Symptom-based control insufficient. Schools may be foci of transmission.	Bell *et al.* 2006, Tellier 2006[28]
Timing of peak shedding	▪ First 1-3 days since symptoms (illness)	More rapid transmission*	Bell *et al.* 2006[55]
Mode of transmission (rank order)	1. Virus-laden large droplet in cough/sneeze 2. Virus-laden small aerosol in cough/sneeze 3. Direct contact with secretions, fomites**	Surgical face masks may not offer complete protection	Tellier 2006
Survivability	▪ Infectious for <24-48 h from steel/plastic ▪ Infectious for <8-12 h from tissue/cloth ▪ Humidity, heat reduce survivability in air	Malls and crowded transport a source of infection	Bell *et al.* 2006, CDC 2006[13]

* In comparison, SARS has peak infectivity 5-10 days after onset of symptoms.
** Any inanimate object that transfers the virus person to person, e.g. dust particles in a sneeze.

Table 1 presents conclusions from recent review studies of the basic infectivity of influenza A viruses, themselves based on the literature on seasonal (inter-pandemic) epidemics in the West. Any deficiencies in this literature are noted in the review studies cited in Table 1. If it is assumed that a pandemic capable virus would have similar properties, then some implications for pandemic control can be gleaned. These will be later put into the context of specific policies for local/community or international preparedness. It can be noted that some aspects of infectivity for the influenza A virus seem higher than the SARS coronavirus from 2003. The caveat should be reinforced that Table 1 is based on historical studies of seasonal influenza A/B, across many different subtypes. Any future pandemic-capable strain may have different properties.

An additional characteristic of infectivity is the presence of age-related patterns in attack rates. Based on data from the US, attack rates were much higher in children than in adults during the 1957-58 (subtype H2N2) pandemic, but were equalized in the 1968-69 (subtype H3N2) pandemic[3]. It is unclear if the 1957-58 pattern was driven by the variation in basic infectivity across age groups (of which there is some evidence vis-à-vis viral shedding in seasonal influenza), by subtype-specific properties, or by the variation across ages in modifiers of basic infectivity such as contact rates. It may be difficult to predict an age pattern of attack rates in a future pandemic, though such a prediction would help setting prevention priorities by age group and locations (e.g., schools vs. offices).

Virulence: The base virulence of a pandemic influenza virus refers to its ability to cause serious health outcomes in an untreated clinical case of the disease. Given virulence, the likelihood of severe outcomes in a case depends on individual level factors (e.g., a complicating comorbidity), and the use and efficacy of pharmaceuticals in treatment. There are several operational measures of basic virulence based on the particular outcome, e.g., deaths per hundred cases or case fatality rate (CFR), hospitalizations per hundred cases, etc. Virulence is directly connected to the ultimate impact on individuals and the economy if an outbreak becomes an epidemic. In theory, virulence is also connected to the duration and attack rate of an epidemic in the no intervention scenario. If a virus incapacitates or kills those infected so rapidly that they do not transmit it to many susceptibles then the epidemic grows slowly. In modeling, this is shown by holding the infectivity constant and increasing the death rate, which usually increases the total duration of a wave of the epidemic, and depending on the model, reduces the total infections[4]. In the intervention scenario, higher virulence may actually assist control as case detection becomes faster and easier[5].

Since virulence during an epidemic of influenza is even more affected by personal characteristics of the individual, it should be discussed in terms of population-level averages. In this context, the projected population-level virulence of a future pandemic could compare to that in prior pandemics or seasonal epidemics. The data from individual human cases of the current avian H5N1 subtype, an overall CFR of about 60%[6], are not as useful, since a pandemic capable virus may trade off some virulence for transmissibility.

Table 2: *Historical Estimates and Modeling Estimates of CFR by scenario*

Scenarios based on history	Lower	Medium	Upper	'Ultra'
Category 5 or Like 1918-20	0.2-0.5	2.5	4	
Category 2 or Like 1957-58	0.04		0.27	
Category 1 or Like 1968-69	0.01	0.013	0.07	
Other modeling estimates				
US, Std Risk 0-19 yrs	0.001	0.002	0.013	Meltzer *et al.*[9]
US, High Risk 0-19 yrs	0.013	0.022	0.765	
US, McKibbin & Sidorenko*	0.023	0.233	1.166	2.333

* Same for all age groups, and run for Mild/Moderate/Severe/Ultra.

The total cases and deaths in the historical pandemics are unknown, but they have been categorized by their estimated CFR, with 'Category 5' being the most devastating[7]. In a study, the ranges of CFR for each historical pandemic were shown as 'lower, medium, and upper' scenarios (Table 2) relevant for a future pandemic[8]. Another study used mortality from seasonal influenza epidemics in the US as a basis for projecting the CFR for a pandemic

in the US[9]. From that study, the modeling estimates of CFR for two risk categories in the 0-19 age group (based on the presence of a complicating comorbidity) are shown in Table 2.

Similar to infectivity, any projected patterns in virulence – across age groups, location, or ethnicity – are important for setting priorities in prevention and mitigation of a pandemic. These patterns are also significant towards the macroeconomic impact. It is uncertain which factor, if any, will cause the definitive pattern in a future pandemic. It is now known that mortality in a case of pandemic influenza A virus of close avian origins – as suspected for 1918-20[10] – results from an inimical feedback process in the immune system, termed a 'cytokine storm'. This response has also been observed in human cases of the current avian H5N1 virus. Here, a healthier immune system, as in younger people, may result in higher risk for severe health outcomes. Still, the age pattern of excess mortality in the 1918-20 pandemic – higher in young adults compared to the elderly – was not repeated in 1957-58 or 1968-69 (these pandemic virus strains had some avian influenza genes). For other patterns, a regression study using international data from 1918-20 found that low per-capita income was a statistically significant explanatory factor of higher mortality[11]. A study of the same pandemic in Iran found significant rural-urban differences in mortality[12].

Effective contact: Given the recurrence of the idea of effective contact in the modeling of influenza transmission, a short discussion is provided. An effective or sufficient contact between an infected and an uninfected, susceptible individual is one able to transmit infection. Depending on the requirements of the transmission model, an effective contact 'rate' (no. of contacts per unit time) or 'probability' (expressing the likelihood of making at least one effective contact per unit time) can be defined. Given a base infectivity of the influenza virus, the effective contact rate/probability varies across individuals based on the volume, frequency, and proximity quotients of their contacts. For example, compared to adults, elementary school children may meet a larger number of contacts in playgrounds and classrooms. Besides the fact that children may shed more virus, the transmission risk per contact may be higher since they engage frequently and get in closer proximity (within three to six feet of each other[13]). However, proximity becomes a less strict criterion in closed, cool areas with minimum airflow, such as an aircraft cabin. In another example of variation, a rural, dispersed population may have lower volume and frequency of contacts.

In modeling studies, effective contact properties are purposively fixed for certain groupings of individuals. The parameter values are often estimated from observational data

using an equation such as $\beta_i = \gamma_i * \rho_i$ (where β is the effective contact rate, γ is the total number of contacts per unit time, and ρ is the risk of transmission given contact, with i as a subscript referring to the particular population group). The estimates of total contact rates across various groups describe a social network. Other aspects of social network analysis are also important for modeling influenza transmission, as will be discussed later.

Concepts in modeling transmission and control of pandemic influenza: The discussion above sets the stage to discuss concepts that in modeling determine whether an infectious disease will be rapidly controllable after the initial emergence of an outbreak:

- The *basic reproduction number R_0* is a measure of the base or intrinsic infectivity of any infectious agent such as an influenza virus. It is defined as the number of secondary infections generated by a primary infection in a homogeneous population where everyone is equally susceptible to infection (complexities in defining R_0 in a heterogeneous population are omitted). R_0 is a modeling construct that captures several characteristics of infectivity – the transmission risk per contact, duration of the infectivity period, mode of transmission, etc. – and is difficult to predict in advance of an epidemic. Some estimates of R_0 for past outbreaks are provided in Table 4.

- The *disease generation time T_g* is the mean interval between infection of one person and infection of the people that the individual infects[14] (Ferguson *et al.* 2005 use T_g=2.6).

- The *proportion of transmission occurring prior to symptoms* (i.e., asymptomatically), θ.

The importance of these constructs in modeling can be seen against modern practice in outbreak control. These public health practices – in the absence of effective vaccine or treatment in the prevention of spread – are credited with containing the spread of SARS. The first actions are active surveillance and then the effective isolation of individuals displaying pre-identified symptoms. The second is the tracing and quarantining of the contacts of the individuals who are displaying symptoms. Success here depends on whether the symptomatic individuals are rapidly identified by authorities (e.g., at airports) or they voluntarily report to a public health facility soon after their symptoms emerge.

Given the importance paid to contact tracing, symptom-based screening, and surveillance in discussions on pandemic influenza, it is useful to review the SARS experience. On the plus side, SARS symptoms preceded the peak infectivity period by a few days. There was little evidence of asymptomatic transmission. This implies that θ was low. The outbreaks also happened to occur in cities with well-functioning public health systems

where trust in official communications and measures was high and maintained. In the negatives, the symptom-based entry screening process at Canadian airports was not found cost-effective[15]. If θ is low but R_0 is high (see Table 3) in a SARS-like outbreak, it may be preferable to rely on surveillance to detect the cases early and then to quarantine quickly.

Table 3: *Estimates of important modeling constructs for infectious disease outbreaks*

Disease Outbreak	R_0	θ	Source
Pandemic influenza, 1918, New Zealand	$1.3 < R_0 < 3.1$	n.a.	Sertsou *et al.* 2006[16]
Pandemic influenza, 1918, UK (first wave)	$1.7 < R_0 < 2$	n.a.	Ferguson *et al.* 2006[37]
Pandemic influenza, 1968-69, Hong Kong	1.89	n.a.	Rvachev & Longini 1985[31]
Smallpox	$4 < R_0 < 10$	$0 < \theta < 20\%$	Eichner & Dietz 2003[17]
SARS 2003 Hong Kong	$2 < R_0 < 4$	<11%	Fraser *et al.* 2004[14]

The higher is R_0 (and lower is T_g) the quicker do public health authorities need to intervene with screening, case-patient isolation, and enhanced active surveillance of the population. The higher is θ, the lower the cost-effectiveness of symptomatic screening. Authorities may not be able to remove infectives fast enough before they cause secondary cases. If a pandemic influenza outbreak occurs, it may be possible to quickly estimate θ from available data on infected people and their contacts. Estimating R_0 is harder, but together with θ, the parameter would set valuable context for the efficacy of proposed policies. Based on what is known of seasonal influenza as well past influenza pandemics, it is believed that:

- Symptomatic screening will be difficult (initial high fever, headache and respiratory symptoms are shared with a few other diseases, including non-pandemic influenza).
- A person with influenza-like symptoms may not self-isolate for up to a day, during which time they would be infective, given what is known about R_0 and θ.
- Even with screening, many infectives would not be detected and isolated (one study found that up to 83% of infectives entering the UK via air travel may be missed[18]).
- R_0 for prior pandemics of influenza was relatively low, lying between 1.5 and 4. Where a future pandemic's R_0 lies in this range has huge bearing on policy efficacy. Certain threshold values for R_0 have been estimated to discuss the efficacy of standard policies. These will be discussed further below when comparing policies.

The efficacy of vaccines in prevention: The US government has allocated 58% of the 2006 pandemic influenza supplemental budget to vaccines (16% on antivirals)[19], perhaps based on success with seasonal influenza. Vaccine efficacy is defined both for reducing susceptibility (VE_s) and in reducing infectiousness (VE_I), both of which prevent an outbreak from growing. Though feasibility and financing problems may prove to be larger

obstacles in the developing world's adoption of pharmaceutical policies, the efficacy issue should be considered. A brief overview of the issues in vaccine efficacy is provided below.

The consensus among experts is that a vaccine based on currently circulating avian influenza A (H5N1) strains that have attacked humans will not prevent infection in most susceptible individuals from a mutated, pandemic-capable strain. However, since these vaccines contain antigens matched to the H5N1 pandemic candidate subtype, there may be immunity against a pandemic strain in some individuals, and more widespread immunity against any currently circulating strain (including those the vaccine does not derive from)[20]. Based on an assumption of partial protection, the vaccines based on current strains are classified as 'pre-pandemic' vaccines. Governments have stocked these in risk-limiting moves. A pre-pandemic vaccine was licensed by the US Food and Drug Administration[21]. Other pre-pandemic vaccines are in stages of development. Without delving into issues of feasibility or ethics (considered later), the following factors modify efficacy:

- Efficacy depends on priorities. Reducing the total number of infections requires vaccination of those most likely to be infectious. Targeting severe health outcomes requires vaccination of those at most risk of severe morbidity or mortality. Optimum policy mixing or a single choice here could depend on the infectivity of the virus, as a study finds[22]. This is considered later below.

- Efficacy of pre-pandemic vaccines is difficult to determine, and can only be reported as immunogenicity (ability to create an immune response, i.e., matched antibodies). Current studies show immunogenicity is dependent on dose size. This is an issue if antigen supplies are limited and if as reported, the dosage for pre-pandemic vaccines without adjuvant is high, e.g. only a 90μg dose reached reasonable immunogenicity in the vaccine licensed by the FDA[23], six times the volume of a standard 15μg seasonal influenza dose. For some vaccines, adjuvant reduces the antigen required[24]. The issue of dosing is considered in more detail later.

Efficacy of chemoprophylaxis: If there will be delay in the availability of an effective vaccine, antiviral prophylaxis has a role. Modeling studies have considered policies of targeted antiviral prophylaxis around an outbreak to prevent its spread. Many countries have stockpiled antiviral doses, including in the developing world. However, there is debate about the efficacy of antivirals in containment, specifically neuraminidase inhibitors. In this context there is confusion over an antiviral formulation's efficacy for primary prevention (pre-exposure prophylaxis of infection), secondary prevention (post-exposure risk

reduction of complication or disease progression), and treatment (therapeutic, to restore health of patients). The quality of evidence varies across these roles for antiviral use in human cases of avian H5N1. Some studies had found oseltamivir – a neuraminidase inhibitor – to be less effective for treatment in such cases, but since these were small-observational rather than clinical trial studies, the WHO has classed the results as weak evidence. Also, it appears that such oseltamivir-resistant avian H5N1 'wild type' viruses may have lower fitness for transmission[25], but with more cases of transmission and antiviral use, compensatory mutations may emerge. If sustained drug resistance does appear during post-exposure prophylaxis, antiviral use would have to cease to prevent further selection of a drug-resistant pandemic virus[5]. This would shatter currently favored mixed intervention strategies. Drug resistance in influenza A will need to be continuously monitored. Considering the debate and the evidence, WHO has recently published guidelines[26] for *post-exposure* chemoprophylaxis of human cases of avian H5N1:

- Oseltamivir (*Tamiflu*) with zanamivir (*Relenza*) as an alternative is strongly recommended for high risk groups, at a seasonal influenza dosage. Amantadine is strongly not recommended. Both recommendations based on low quality evidence.
- Oseltamivir with zanamivir as alternative is weakly *not* recommended for moderate and low risk groups and *strongly not* recommended for low-risk pregnant women.

There are no clinical studies of antivirals in the primary preventive role for human cases of avian H5N1, or for their efficacy in reducing the infectiousness of exposed individuals. This must be kept in mind when targeted antiviral prophylaxis policies are discussed below. A study[27] proposes the relation $R_{av}=R_0/3.6$ where R_{av} obtains with preventive, mass, and prolonged oseltamivir prophylaxis. In the absence of evidence, any reading of the *post-exposure* chemoprophylaxis guidelines above as indicative of efficacy in reducing susceptibility or infectivity with antivirals should be undertaken with caution.

Efficacy of physical prophylaxis: Vaccines and antivirals may theoretically limit both susceptibility and infectiousness, analogous respectively to ex-ante prevention and ex-post mitigation of outbreaks. Similarly, personal protective equipment such as N95 or higher *respirators* (rather than surgical facemasks) could prevent a person getting infected if they are susceptible or from infecting others. The availability of such respirators is likely to be limited for community outbreak control even in industrialized countries. Other equipment, such as gloves, is designed to prevent an uninfected person from coming in

contact with fomites and respiratory secretions. During SARS, it was common to see individuals undertake self-protective behavior by wearing facemasks.

However, if influenza is also transmitted in aerosols emitted in sneezes of coughs, i.e. particles of diameter less than 10μm (in comparison, large droplets are defined as having a diameter of 50-100μm), then surgical masks may not offer effective prophylaxis. Some review studies have indicated that aerosol generation is likely in the average individual with influenza, and this is increased by aerosol-generating procedures in healthcare settings (e.g., endotracheal intubation, open suctioning)[13]. Besides the fear of them slipping through the average facemask, aerosols are a concern since they may stay airborne longer[28]. However, they may also carry a smaller infectious dose. More research is required on these characteristics, especially as the efficacy of facemasks, which are affordable enough for personal use and local stocking, is of prime importance in community outbreak control. At present this efficacy remains controversial. Some recent laboratory studies using test aerosols of appropriate diameter (but not infectious agents) have shown that facemasks may have protective properties comparable to high end respirators[29, 30].

Predicting the community transmission of pandemic influenza: an overview of methodologies

Epidemiological modeling of future pandemic influenza transmission is complex because the biology of the virus, responses of a population, etc., are difficult to predict beforehand. Non-epidemiological studies from the impact literature use historical data from 20th century pandemics to estimate a flat rate of infections across the entire population during a wave, known as the Gross Attack Rate (GAR). These assume the same risk of infection for every individual, and hence GAR is applied as a fraction of the population. The section below discusses methods for the *epidemiological* modeling of infections in a wave of pandemic flu transmission. Results of the studies are compared in the next section.

Compartment or SIR/SEIR models of pandemic flu transmission: The simplest non-stochastic (deterministic) models of influenza transmission - pandemic or seasonal – are based on SIR (Susceptible, Infected, and Recovered) models of 1920s vintage. By adding the 'exposed' category, SEIR models are the root of *deterministic modeling* of epidemics. Fig. 1 on this page is an unorthodox representation[59], modified to show some (but not all)

impact points of public health interventions. Individuals transition based on predetermined probabilities from being susceptible (S), to exposed (E, i.e., infection is latent), to infectious (I) and then to removed (R, recovered or dead). Other preset parameters are the length of time individuals stay in state E without becoming infectious (latency period), and the time they stay infectious before being removed. The model is specified in differential equations that define the stock of people in the various states after a discrete time step (usually a day). At time zero, a number of seed infections are introduced and the model is run.

Fig. 1: Simplified SEIR model diagram with potential interventions
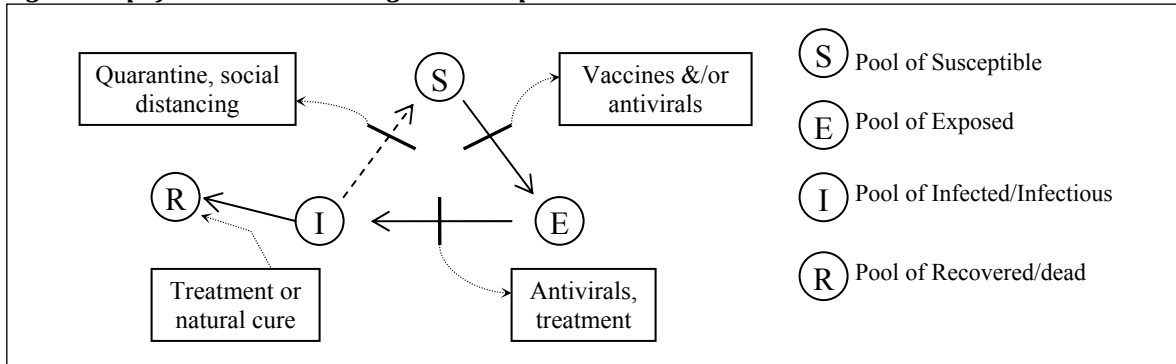
For pandemic influenza, the possibility that any infected person contacts a susceptible during a time step is set as the fraction of the population susceptible at that point, multiplied by a fixed decimal: the 'probability of infection conditional on contact', analogous to transmission risk as introduced earlier[31]. The method has the assumption of population homogeneity with 'uniform mixing'. This implies every person is equally at risk of infection and every individual in the population is equally likely to contact someone else.

The simple SEIR models neglect the fact that individuals do not mix uniformly; each person has contact with only a small fraction of a population. Also, the population is not homogenous – some individuals have more contacts and meet them more often. Therefore, while the first real world factor reduces the risk of infection for the average individual, the second may raise it for some individuals, for example, urban citizens vs. those from rural, low population density areas. In reality, risk of infection varies widely across people.

The policy intent of SEIR models is to describe transmission dynamics at the population level by varying attributes of individuals that locate them within compartments (disease state, age, reported case or hospitalization status). These models may predict natural quenching of an epidemic (i.e., due to 'population/herd immunity' when the number

of susceptibles left falls below a sustaining threshold[32]). Public health interventions – such as antiviral prophylaxis – can be introduced, and the effect observed via a change in the probability of transmission per contact across the various infective group members and the remaining susceptible. For capturing more of the variation in risk of infection across individuals, SEIR models have to add compartments without losing mathematical tractability. Several improvements in this vein have been made, which can be discussed by relating them to the effective reproductive number of the epidemic, or R.

In this context, the basic reproductive number, R_0, captures the *a priori* infectiousness of a communicable disease in a particular setting. It is defined as the number of secondary infections caused by a single typical infected case in a completely susceptible population, and in the absence of interventions. At the start of an epidemic, it determines how quickly the epidemic will spread. The effective reproductive number, R, is equal to R_0*s where s is the proportion of the at-risk population still susceptible. Theoretically, if R can be pushed below 1, an epidemic usually dies out due to herd immunity[32]. Compared to the *a priori* R_0, R is always lower. Given any initial 'intrinsic' R_0, R for the epidemic depends on T_g, or more generally on factors including:

- The risk of transmission per proximate contact, ρ
- The number of effective contacts an average person has per unit time, β
- Relative duration of the latent and infectivity periods

As briefly discussed previously, R is an important factor[33] in determining how many infections will be suffered in one wave of the epidemic, i.e., the gross attack rate. The widely reported 'threshold R_0' effects for the efficacy of certain interventions should be seen as model-specific and indicative rather than definitive for policy analysis. Early and pervasive public health communications after an outbreak of pandemic influenza can reduce the average R over any subsequent epidemic wave (which may be prevented entirely by successful outbreak containment). For example, by forcing or urging people to limit contacts, encouraging hand-washing and other personal hygiene, or promoting the use of facemasks. Historically, public health communications became effective enough to modify overall R during subsequent epidemic waves when the authorities were poised to begin them early, and similarly in cities and countries with later outbreaks[34,56].

Fig. 2: *Theorized relationship between R_0, number of people infected, and epidemic duration*
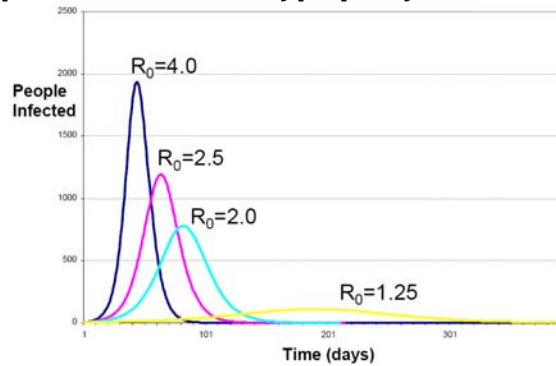
Source: Risk Management Solutions (2007)

Population heterogeneity and SEIR: In the *no intervention* case, actual *R* may vary across population groups based on age structure and the types of social mixing situations. The SEIR models can be improved to account for age structure. The importance of age as a determinant of susceptibility to a clinical case of pandemic influenza infection may be due to both behavioral and biological factors. The W-shaped shaped curve relating excess mortality rate in percent during the 1918 pandemic (y-axis) to age groups (x-axis) suggested that deaths were highest in the segments between 15-39 years and lower for higher and lower age groups[11]. Mortality reflects both the underlying attack rate and the case fatality rate (CFR). There are hypotheses for the higher CFR in younger individuals, which relates to hyper-interaction of the symptoms of avian-derived pandemic influenza viruses and the more robust immune systems as present in younger persons. For this study, the implication of the W-shaped curve for the variation in the attack rate is stressed. The portion of the variation in mortality by age during 1918-19 that is explained by age-differed attack rates is unknown[11]. Estimates of age-differenced attack rates using data from later pandemics suggest that a biological factor may be in play, related to the particular influenza A subtype that causes the pandemic. The 1957-58 pandemic with subtype H2N2 displayed a higher attack rate in U.S. children, whereas the 1968-69 influenza pandemic with subtype H3N2 had a similar attack rate across age groups[3]. In general, younger people shed more viral material[15]. As such, the risks of infection for children in schools would be higher.

Behavior also differs across age groups in ways that affects susceptibility. In high population density areas, younger people have contacts more frequently and at closer quarters – e.g., on urban transport, in offices, and on playgrounds. Older people who are

22

homebound and do not have as many contacts have a lower risk of being near an infectious person, but when near such a person, may have a higher risk of getting infected because of weaker immune systems. In SEIR models, one way to simulate these differences in behavioral and biological factors of susceptibility is to introduce varying transmission probabilities defined in advance for different age compartments. With more computing power available in recent years, such age-structured models have become more common.

When simulating the effectiveness of public health policies, as much heterogeneity in risk of infection is desirable as is reasonable to model. Location is another important source of such variation. Urban areas can be quarantined or social distancing measures imposed, which cut the contact rates in the population. But these are more difficult to impose in rural areas where there are less defined modes of entry and exit and the population is harder to reach with public health communication. Therefore, while in rural areas the average number of effective contacts per person might be low, the public health measures might have low impact. And as occurs for older people, rural people on average may have poorer health status. They may also have poor access to antivirals. In the pandemic influenza case, since the entire country is susceptible to pandemic influenza and rapid transmission via air and road networks is probable, modeling simultaneous SEIR progress in multiple cities that exchange susceptible and infectious people is desirable. However, this requires mainframe computing resources with the time step of a day.

Physical distribution and structure of populations: In comparison with non-spatial SEIR, *stochastic-spatial* models have more success in incorporating population and location heterogeneity. They also have a different intent and modeling philosophy. A stochastic-spatial epidemic model simulates the epidemic process as a series of random events in space and time with the probability of specific events defined by the model parameters. The two event probabilities of interest are the probability of effective contact (c), and probability of transmission on contact (x). The former varies at the individual level based on factors – location in space, and characteristics of that location. The latter varies in reality based on age, public health context, individual behavior, etc., and is usually a constant in analysis as it is difficult to model.

In a recent stochastic-spatial model of pandemic flu spread in rural SE Asia the location contexts are called 'mixing groups' (e.g., offices, schools, and households), and for each such context a specific probability of effective contact per day is defined[3]. The age-

distribution of individuals, number of mixing groups and their average sizes, as well as their clustering are set so as to mimic a real population (in this case rural Thailand). Average inter-mixing group distances –important to generate the spread across the simulated space – are based on GIS data. At the time of model generation, 500,000 individuals are distributed according to an algorithm in close contact mixing groups (households, schools, and workplaces), as well as casual/social contact mixing groups (temples, markets, shops). The people have effective contacts based on the assumed average rates for that mixing group context. For example, the authors assume that the probability of two children making at least one effective contact per day in a household mixing group is 0.6[35].

For the base case, stochasticity is introduced by conducting Bernoulli trials and generating N uniform [0, 1] numbers for N mixing groups. If the probability for an individual to be infected on a day in that group[36] is greater than the random number, then a single infected individual in his/her latent period is 'introduced' into each of N mixing groups at model inception. Secondary infections result in each mixing group that has an infected individual. Similar to 'hierarchical epidemic' models, this within mixing group dynamic is extended to between-group analysis. The spread to mixing groups without primary introductions is modeled, based on assumptions and distance from infected groups. Each stochastic realization, for a certain R_0, leads to an overall attack rate.

Such stochastic-spatial models are computationally intensive if populated with millions of individuals, but can evaluate public health interventions such as quarantines and social distancing, ring prophylaxis (where a fixed percentage of people in an x square mile radius around an outbreak are given antivirals), and vaccination. Similar models, but with different assumptions and techniques, have also been constructed for Thailand[5], the US and UK[37], and separately for the US[38]. Important differences obtain in whether the model is capped at producing an overall attack rate for a level of R_0 and in how higher R_0 is reflected.

Stochastic-spatial models capture population-level effects by describing the intensity of risk at the individual level. They can also be calibrated to a specific magnitude of GAR from prior pandemics (Longini *et al.* chose to calibrate to 33%, based on the first wave of the 1957 and 1968 pandemics) and a relative attack rate pattern across ages (see Fig. 3 below). As suggested above, the stochastic model produces a distribution of age-specific and gross attack rates at different R_0. By selecting for an overall GAR of 33% and an age-specific

attack rate pattern, Longini *et al.* determine the model seeding that reproduces past pandemics as a base case. They test the intervention strategies thereafter at this setting.

Fig. 3: *Assumed pattern of attack rates across age groups at different R_0 (Longini et al. 2005)*
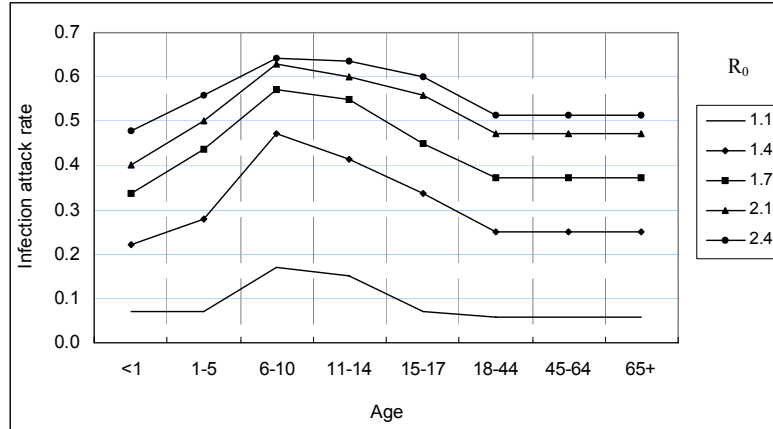
*Pattern is based on US data from 1957-58 (more children infected) and 1968 (similar attack rates across ages) pandemics, mixed with seasonal influenza data from SE Asia (similar to 1957-58 pandemic in the US).

The calibrated stochastic-spatial technique ensures that the locational and demographic heterogeneity of the epidemic area is captured, while the full range of the stochastic results are capped at past experience. This method is good at predicting whether an epidemic that begins randomly somewhere will spread, and evaluating which control options will be successful in limiting its spread.

Table 5 provides a comparison of a model like in Longini *et al.* (2005) and the age-structured SEIR model portion from a proposed World Bank study. The two models differ in the captured heterogeneity in risk of infection and in the intent. Since the pandemic in 1918 had unique features leading to a severe impact, when constructing a model for a high risk Asian country it is important to base it on Asian experience in 1957 and 1968 (as in Longini *et al.* for SE Asia). In general, some parameters are exogenously imposed constants not varied in modeling. Varying parameters allow sensitivity to the modeling assumptions to be tested, necessary given the uncertainty in estimates, e.g., of contact rates. The model should also be portable, with important variables generated endogenously rather than assumed.

Table 4: *Comparison of two model archetypes for predicting pandemic influenza transmission*

Characteristic	Deterministic SEIR: age-structure	Stochastic –spatial model
Source study	Part of a proposed WB study	Longini *et al.* (2005)
Contact rates by age	Exogenous constants#	Exogenous constants
Attack rates by age	Endogenously derived	Endogenously derived§
Contact rates by location	Exogenous constants#	Exogenous constants
Attack rates by location	Endogenously derived	Endogenously derived
Whole epidemic R_0	Not modeled	Exogenously varied

Whole epidemic GAR	Endogenously derived	Endogenously derived
Raw data needs	Demographic data, seasonal influenza contact rate estimates	Demographic data, geographic data, contact rate estimates
Attack rates by policy scenario	Proposed to be modeled	Endogenously derived
Primary policy use	Describe population-level dynamics, estimate of no. of infections	Describe individual-level dynamics, test public health measures

[#] Proposed to create separate matrices of child-child, adult-child, etc., *effective* contact rates for cities by seasons.
[§] Calibrated to fit patterns of age-structured attack rates from prior pandemics at various R_0.

Mathematical or static scenario models of pandemic influenza: Table 4 suggests that existing pandemic flu transmission model archetypes suit different policy questions. If interest is limited to impact analysis of the base case of a pandemic, then complex deterministic or stochastic transmission models are not required. This logic drives recent non-epidemiological impact analyses, with scenarios based on 1918[8]. However, some quasi-epidemiological detail is required if the scenarios are to be more realistic. For resource allocation decisions, the impact estimates should be as contextual as is feasible.

Two studies are briefly discussed that use a mathematical or static scenario-based model to inform domestic policy decisions. Meltzer *et al.*[9] use a Monte Carlo 'mathematical simulation' model to estimate the impact in the US of a pandemic of influenza and scenarios of related key interventions. A similar model is used by Doyle *et al.*[39] for France. The GAR[40] is varied in 5 percentage point increments from 15% to 35%. The total numbers of cases at any GAR are then distributed across age groups in two pattern distribution scenarios:

Age group	Pattern A	Pattern B
	% of all cases	% of all cases
0-19 years old	40	46
20-64 yrs old	53.1	46.7
65+ yrs old	6.8	7.3

Based on upper and lower estimates of age-specific attack rates from 1918, 1928-29, and 1957 influenza epidemics in the U.S (1918 and 1957 were pandemics).

The three age groups also differ in the proportions in each with pre-existing conditions that cause complications once an individual has influenza. These complications are health outcomes such as severe illness and death, and rates of these per age group are distributions with upper/lower limits based on statistics from past influenza epidemics in the US. The result of the study is not a single estimate of the number of clinical illnesses or a dollar figure, but a range for such variables. The authors describe their intent as "altering a number of variables and evaluating how the results affect key (policy) decisions.[41]"

In van Genugten *et al.*[42], researchers using a 'static' scenario method to test policies for pandemic influenza in the Netherlands. The GAR was exogenously set at 30%, and an

age distribution of the cases was obtained by applying a pattern based on seasonal influenza in the Netherlands. The 'scenarios' are separate policy interventions that reduce adverse health outcomes. Compared to Meltzer *et al.*, the study is static since the parameters for health outcomes are all point estimates from prior Dutch influenza epidemics.

Modeling the effectiveness of policies for containing pandemic influenza

As abstractions of reality, models of future pandemic transmission and its control are selective over the range of included policies and the efficacy for each policy. This is important for model parsimony, but the choices for inclusion and exclusion affect the conclusions. The choice set of policies to include for testing in simulations are those suggested by the modeling of potential bio-terror attacks in the US, the experience with SARS, and public health practice over the latter half of the 20th century. The parameter ranges for the efficacy of vaccination and prophylaxis are either assumed in such modeling, or derived from studies of closely linked disease and epidemic situations. Table 5 presents a categorized menu of control policy choices available for a potential pandemic of influenza. Italicized prevention policies have received more research attention, or have had their effectiveness modeled. These are the only policies considered in detail in the section below.

Table 5: *A taxonomy of policies for containment and mitigation of pandemic influenza*

Type	Sub-Category	Individual policies (M: mitigation; C: containment)
Nonpharmaceutical	International	• International land border quarantines (C) • *Air travel restrictions & advisories* (C) • Quarantine of ships & other restrictions (C)
	National or central	• *Inter-city movement restrictions* (C, M) • Aspects of quarantine (C, M) • Public communications (C, M)
	Community or local	• *Case detection and isolation* (C, M) • *Social distancing* (C, M) • *Local quarantines* (C, M) • Public hygiene and disinfection (C, M) • Local public communication (C, M) • Personal protective equipment (C, M)
Pharmaceutical (or drug-based)	Vaccination	• *Targeted vaccination policies* (M) • *Broad-based vaccination policy* (C, M)
	Antiviral prophylaxis	• *Targeted (ring) prophylaxis around an outbreak* (C) • *Prophylaxis based on contact tracing* (C,M) • *Mass prophylaxis in the at-risk population* (M)

Distinctions between C and M only indicate emphasis, as containment achieves mitigation (lowers the GAR).

Based on results from modeling studies reviewed further below, different configurations of policies from Table 5 are currently proposed for pandemic influenza.

These configurations are an example of prospective modeling evidence on effectiveness being used in policymaking. The review of studies follows after an introduction to the field.

In the earliest instance, a group of researchers within the MIDAS (Models of Infectious Disease Agents Study) network in the US proposed a framework for mitigation policies in the US after the emergence of pandemic influenza. The guiding assumptions were of limited antiviral supply with no effective vaccine immediately available, and that interventions must begin rapidly at the *community level* to prevent an uncontrollable spread that overwhelms response. The proposed strategy of 'targeted layered containment' (TLC) attempts to benefit from synergies across the combined policies[43]. The policies considered high priority are: targeted antiviral treatment and isolation of cases, targeted prophylaxis and quarantine of household contacts of index cases, closure of schools and keeping children at home for the duration of the policy, social distancing at the workplace (telecommuting) and in the community (cancelling public events). Since the assumptions on pharmaceutical availability are very relevant for developing countries, this selection of policies from the menu has been influential for modeling and agenda-setting.

The WHO Interim Protocol for community/local control measures[1] sets forward guidelines for the rapid containment of an outbreak of pandemic influenza, differentiating it from the response to current outbreaks of avian influenza. The Interim Protocol samples policies that also appear in TLC above, articulated here for the containment objective. The current version (dated May 2007) takes a geographically based approach where the initial area of the outbreak becomes the main target – the *containment zone* – in which actions are taken to stamp out the infection and prevent its spread. Within the containment zone and in an area around it called the *buffer zone*, surveillance and community mobilization will check and maintain containment. In the buffer zone, any 'break through' cases are quickly detected and isolated. The boundaries of the two zones, the duration of the operation, and the exact choice and intensity of policies are to be driven by local context and the outbreak characteristics. The generally recommended policies in the containment zone follow TLC:

- Antiviral drugs for treatment and prophylaxis
- Restrictions (e.g., screening) on movements within, into, and out of the zone
- Isolation of ill persons, voluntary quarantine for exposed, and social distancing

The sub-section below discusses some modeling studies comparing policy configurations that appear in the TLC or WHO frameworks. Studies with a containment focus influenced the Interim Protocol, but WHO notes these caveats about the models used:

1. Emergence of the pandemic virus is in a localized and circumscribed area
2. The efficient and sustained human transmission of the virus is rapidly detected and reported such that an appropriate containment strategy can begin
3. Availability of minimum drug stockpiles and the ability to rapidly distribute
4. Movement restrictions and other nonpharmaceutical interventions are feasible

Caveats #2-4 involve 'likelihood of use' and feasibility. Most of the studies do note that these issues will be important. However, such realism is difficult to account for in modeling without introducing additional complexity and uncertainty. The comparative studies below allow for the effect of delay in starting policies and find that it reduces the effectiveness for control. This is discussed more specifically below. The models assume baseline efficacy for some policies. Such assumptions on efficacy for a study are noted below. The discussion should be read in light of the prior review of methodologies; the comparative studies all fall in the category of spatial-stochastic models. After the sub-section below, specific studies of a few italicized policies from Table 5 are discussed.

Comparative Modeling of Policies

1. **Longini *et al.* (2005)**[3]: Rural South-East Asia

Policies compared: Various levels of targeted antiviral prophylaxis (TAP), geographically targeted antiviral prophylaxis or 'ring prophylaxis' (GTAP), quarantine, and pre-vaccination; used singly or in combination. The TAP policy is defined here as the treatment of an assumed percentage of identified index cases (i.e., the first symptomatic illness in a particular mixing group – mixing groups defined in the previous section), and prophylaxis for all their close contacts if the case belongs to a household or preschool group, and some assumed percentage of their contacts if in a workplace or other school group. The antiviral used is oseltamivir, and a single course is administered to the contacts at the same time that therapeutic treatment of the index cases begins (assumed 1 day after symptoms). The GTAP policy is defined on the understanding that identifying index cases spread across 'mixing groups' will be resource intensive. In GTAP, a geographical approach is taken, and once an index case is identified in a locality, an assumed percentage of the people in the entire locality are given one course of oseltamivir. Quarantine is also defined at the locality level: once an index case is identified, all infected plus an assumed percentage of susceptible restrict their movements to the household or neighborhood. Pre-vaccination

occurs before the pandemic and those vaccinated develop some level of immunity. All the italicized percentage levels of the policies are referenced in Table 7 below.

Assumed pharmaceutical efficacy: Vaccine efficacy for susceptibility (VES) is assumed to be 0.3, and for infectiousness (VEI), 0.544. These are low values. Oseltamivir efficacy for susceptibility to infection or primary prevention (AVES) is assumed to be 0.3, efficacy for secondary prevention (AVED, against disease progression) is assumed to be 0.6, and efficacy in therapy for symptomatic disease is = 1 – (1- AVES)(1- AVED), equal to 0.72. Oseltamivir efficacy for infectiousness (AVEI) is assumed to be 0.62. The values for antiviral efficacy are derived from prior studies of household infection with seasonal influenza[45].

Control objective: Containment of cases (i.e., symptomatic infection) at ≤ 1 per 1000.

Table 6: *Effectiveness of policies for control in a population of 500,000* *(Longini et al.)*

Percentage application of policy	Cases per 1000		Containment proportion#	
	$R_0 = 1.4$	$R_0 = 1.7$	$R_0 = 1.4$	$R_0 = 1.7$
a. No intervention	211	384	-	-
b. 80% TAP	0.13	149	98%	33%
c. 90% GTAP	0.28	54	95%	59%
d. 80% TAP + 50% pre-vaccination	0.02	0.16	100%	98%
e. 80% TAP + 70% pre-vaccination	0.01	0.04	100%	100%
f. 70% quarantine	0.17	1	98%	57%
g. 80% TAP + 70% quarantine	0.06	0.14	100%	100%
h. 80% TAP + 70% quarantine + 50% pre-vaccination	0.02	0.03	100%	100%

Proportion of simulations with the policy in which cases per 1000 are ≤ 1

Discussion: The results in Table 6 show a dependence on the value of R_0, especially for the effectiveness of TAP, GTAP, and quarantine. As R_0 increases from 1.4 to 1.7, the required courses of antiviral in TAP (not shown) increases 78 times. Mixed strategies are more robust (***d-e, g-h***). The authors expect initial R_0 for an emergent influenza strain to be below 2, which implies a role for antivirals at their assumed efficacy (the results are moderately sensitive to the assumed AVE$_S$ value, as may be expected). This finding has led to recognition in the policy sphere of 'threshold R_0' effects for policy effectiveness, i.e., with a more infectious viral strain, containment may be very difficult. As R_0 may increase during transmission as the viral strain gains fitness through mutation, the authors suggest early intervention is very important. This insight is valuable.

A caveat is placed on the results. The dependence on R_0 is in part a model artifact. Longini *et al.* calibrate the baseline model at R_0=1.4, which fits the historically observed overall 33% attack rate, and further to an age-specific attack rate pattern as in Fig. 3. They then generate the other R_0 levels (e.g., R_0=1.7) by increasing the transmission probability ρ

per contact, fixed across all age groups and contexts. Recall the equation discussed above for the effective contact rate, $\beta = \gamma * \rho$ (where γ is the total number of contacts, and ρ is the transmission probability per contact). In the Longini *et al.* model, higher R_0 mechanically generates more cases from the channel of ρ. Hence, single effect policies which only impact ρ fare poorly from increases in R_0, but mixed policies (ρ and γ) are robust.

In a real epidemic, there would be differences, but deviation from the results in Table 6 is uncertain. Targeted antiviral use may raise awareness and lower willingness to mingle, hence affecting γ, besides the effects on ρ[46]. Additionally, public communication, low cost yet not modeled, could cut R levels by encouraging personal protective behavior.

2. **Ferguson *et al.* (2005)**[5]: Thailand plus 100-km wide zone of contiguous border nations

Policies compared: The policies considered are similar to Longini *et al*. Social targeting of antivirals (TAP variant) is considered for a percentage of pupils or colleagues in an assumed proportion of the schools/workplaces with index cases. So is GTAP (here called ring prophylaxis), except it is defined as the prophylaxis of the entire population within a ring of a certain radius (5, 10 or 15km) centered on each index case. A drug-sparing variant of ring prophylaxis is considered, where only the nearest 10-50,000 people within 10km of an index case are given a course of antivirals. For nonpharmaceutical measures, they consider social distancing (school and workplace closure), but allow that these might have unforeseen effects by increasing household and random contact rates by 100% and 50% respectively. They also consider an area quarantine policy. Specific definitions of the two latter policies are given along with Table 7 below, which shows the results for selected values of the policy implementation levels. In this study the authors exclude vaccination.

Assumed pharmaceutical efficacy: Ferguson *et al*. (2005) assume oseltamivir efficacy for susceptibility to infection to be 30%, and efficacy for secondary prevention at 65%. Oseltamivir efficacy for infectiousness is assumed to be 60%. Over the entire course of treatment, oseltamivir reduces total infectiousness by a maximum of 28%. The assumptions reference work on resistance to oseltamivir, and other studies.

Control objective: Containment: increasing the probability of eliminating a large epidemic. The 'large' criteria is likely to be geographical, i.e., a country-wide epidemic.

Table 7: *Effectiveness of policies for control in a population of 85 million* (Ferguson et al. 2005)

Percentage application of policy	Av. no. of courses (10^6)		Prob. of elimination (%)	
	$R_0 = 1.5$	$R_0 = 1.7$	$R_0 = 1.5$	$R_0 = 1.7$
1. 90% TAP + 90% GTAP[ξ] in 90% of cases	≈1.8	≈2	≈92%	**0%**
2. Drug-sparing GTAP[*]	≈1.2	≈1.6	≈95%	≈75%
3. Drug-sparing GTAP[*], 80% quarantine[**]	≈0.75	≈1.25	100%	≈92%
4. Drug-sparing GTAP[*], social distancing[#], 80% quarantine	≈0.75	≈1.4	100%	≈99%
5. 90% GTAP[ξ] & 80% quarantine, but with a 3 million courses limit	-	-	≈92%	≈80%
6. Drug-sparing GTAP[*] & 80% quarantine, but with a *1* million courses limit	-	-	≈90%	≈60%

[ξ] Within a 5km radius of an index case   [*] 50,000 courses (people) within a 10km radius of an index case
[#] 80% reduction of movement in and out of a zone defined by merging the 5km rings around index cases
[**] 21-day closure of 90% of schools and 50% of workplaces within 5km of an index case

Discussion: Table 7 presents a comparison of outcome measures for selected levels of implementation (mostly the median) for policies at levels of R_0. In their paper, the authors represent outcomes graphically, and allow for 95% confidence limits for results at a particular R_0 and level of implementation. Therefore the ≈ sign in Table 7 represents approximation for discussion purposes.

The threshold effects of R_0 on policy effectiveness seen in the Longini *et al.*[3] study have intensified. A pure targeted antiviral policy (TAP and GTAP) does not provide containment at higher R_0. Mixed pharmaceutical plus nonpharmaceutical policies (e.g., **4** in Table 7) achieve full elimination at lower R_0, and up to 90% at R_0=1.9, a trend similar to Longini *et al*. Containment with TAP+GTAP fails at lower levels of R_0 if delays in policy initiation grow from 0 to 4 days (not shown in Table 7). It is not clear why a drug-sparing GTAP policy alone performs better than TAP+GTAP. Even at 10,000 courses per index case (lower than the 50,000 courses case in Table 7), the containment performance is comparable to TAP plus a 5km GTAP policy. The results may reflect the choice of initial outbreak seeding: a sparsely populated rural area, where a prophylaxis policy of 10,000 courses buys more coverage than ring prophylaxis of radii 5-10km.

The authors show that a low stockpile of antivirals could be significant constraint for outbreak containment at higher levels of R_0 in a large population. For a modeled Thai population of about 85 million – without considering the neighboring countries' border areas – a stockpile of 3 million+ provides reasonable prevention if mixed strategies are considered. The authors find the results dependent on the assumptions of antiviral efficacy. Low sensitivity (e.g., picking up less than 40% of infection) of case detection impacts containment, while low specificity (false positives) wastes drugs and logistical capacity.

3. **Ferguson *et al.* (2006)**[37]: The United States and Great Britain (GB)

Policies compared: The policies compared here derive from Ferguson *et al.* (2005), but with assumed implementation level and efficacy suited to the context of a country at risk of secondary outbreaks after initial occurrences elsewhere globally; and where subsequent internal spread is likely to be rapid[47]. Social targeting of antivirals (TAP variant) is considered for a percentage of schoolmates or work colleagues of the index cases. GTAP is excluded, but a pre-vaccination policy is introduced. Here various nonpharmaceutical measures are key. The possibility that outbreaks will occur in the developing world means that entry restrictions air and border control) may delay and/or limit the outbreaks in the US and GB. Various types of movement restrictions are considered – reactive (RMR, where a 20km exclusion zone is established around every index case and movement in and out is stopped), and blanket (BMR, where journeys over 20-50km from the home are stopped). They consider social distancing (school and workplace closure), but allow that these might increase household contact rates by 100%. They also consider a quarantine policy for households with index cases, with a 50% compliance rate. Early case detection and treatment is evaluated, but note that the assumption is that *only 50%* of symptomatic illnesses are reported or targeted (across policies **5-25** in Table 8). Results for the US at select implementation levels of various policies (and related specifics) are shown in Table 8.

Assumed pharmaceutical efficacy: Assumed oseltamivir efficacy is as in Ferguson *et al.* (2005), except reduction in total infectiousness – given time delays in detection and treatment– is capped at 25%. For a pre-pandemic vaccine, assumed efficacy for reducing susceptibility to infection is 30%, efficacy for secondary prevention is 50%, and efficacy for reducing infectiousness is 30%. For an assumed *pandemic* vaccine, these values are 70%, 50%, and 30%. A single dose is assumed sufficient, and vaccine coverage is set at 90%.

Control objective: Mitigation. Outbreaks in the US/GB are considered inevitable and mitigation policies begin with a delay from the global emergence. Policies are compared on the delay in the US/GB epidemic peak and the reduction in total cases that they achieve.

Table 8: *Effectiveness of policies for control in the US pop. of 300 million (Ferguson et al. 2006)*

Percentage application/effectiveness of policy	*Delay in US peak (days)*		*Cum. attack rate %*	
	R_0=1.7	R_0=2.0	R_0=1.7	R_0=2.0
1. No intervention	0	0	27%	34%
Entry control and movement restriction				
2. 90% effective border control[ξ]	≈15	≈10	-	**-**
3. 99.9% effective border control[ξ] + air restrictions[*]	≈50	≈42	-	-
4. 99.9% effective border control[ξ], BMR 20km[**]	≈60	≈52	-	-

Case detection, treatment, and isolation				
5. Same day treatment for all reported cases[***]	-	-	≈25%	≈31%
6. As **5** but 90% receive, and with 2 day delay[***]	-	-	≈25%	≈31%
7. As **6** but same day treatment[***]	-	-	≈22%	≈29%
8. Same day case isolation, 70% of cases[#]	-	-	≈21%	≈28%
9. Same day case isolation, 90% of cases[#]	-	-	≈18%	≈25%
Household policies & prophylaxis policies				
10. Antiviral treatment for 90% cases + prophylaxis for their household contacts[¶]	8	6	17%	22%
11. As **10** plus prophylaxis of school/work contacts	18	26	7.6%	13%
12. 14-day quarantine for households with a case[¶¶]	4	2	24%	30%
13. Combination of **10** and **12**	10	8	15%	20%
Social distancing policies (school/workplace)				
14. 100% reactive school & 10% workplace closure[§]	13	9	24%	32%
15. As **14** but with 50% workplace closure	16	11	23%	31%
16. Mass vaccination from day 30 of world outbreak beginning with 0-16 y/o[¤]			0%	1%
17. As **16**, but with start from day 60 of world outbreak[¤]			19%	31%
18. Random vaccination from day 60 of world outbreak[¤]			4%	16%
19. As **17**, but beginning with those over 60 y/o[¤]			21%	31%
Profiled mixed strategies (as in Ferguson et al. 2006)				
20. Household quarantine[¶¶] + 100% reactive school closure[§]			21%	29%
21. As **20**, plus 50% next day case treatment			19%	27%
22. 100% reactive school closure plus 90% next day case treatment			22%	30%
23. As **22** plus household prophylaxis as in **10**			13%	20%
24. As **23**, plus *pre-vaccination* of 20% of population, prioritizing 0-16 y/o			7%	14%
25. As **23**, plus 80% prophylaxis of school/work contacts, plus 99% effective border controls			1%	10%

[ξ] Targeted at inbound intl. travelers (air and land)   [*] Full closure of domestic air traffic
[**] Journeys over 20km from home are banned   [***] Delay between symptom onset and antiviral treatment
[#] Assumed to cause 90% reduction in contacts   [¶] Prophylaxis with delay of 1 day since symptoms of the case
[¶¶] With 50% compliance   [§] Only a % of schools/workplaces closed that have a detected case (reactive)
[¤] Mass vaccination at the rate of 1% of the population per day since inception of policy

Discussion: The authors try policies for the prevention of a large caseload in the US given that outbreaks are *inevitable* (a specific critique of this follows further below). All policies **2-25** are sensitive to R_0 and policies **5-9** and **16-19** are sensitive to time delays. Even prompt and pervasive case detection (policy **9**) or treatment (policy **6**) does not offer substantial containment. An approximate 18% attack rate in the former will still be a major health event for the US. An intensive antiviral prophylaxis policy (**11**) will substantially contain the attack rate, but will require a stockpile sufficient to cover 72% of the population at R_0=1.7 (not shown in Table 8)[48]. Based on the delayed availability of a *pandemic* vaccine (and assumed efficacy as noted above), substantial containment can be achieved if a mass vaccination program begins reasonably quickly and is targeted at children (policy **16**). However, the 30-day timeline (since a circulating pandemic virus strain is identified) for the production and mass deployment of a pandemic vaccine is unlikely. A random vaccination policy (**18**), beginning after 60 days since the start of the world outbreak performs better

than age-specific vaccination policies. These results assume a single dose is sufficient. If two doses – as some vaccines require – must be administered one month apart, then a pandemic vaccine must be ready for distribution at 30 days.

These results point to the need for mixed strategies. The mixed strategies here do not involve a pandemic vaccine policy. Of all the potential combinations (the authors provide about 30 mixed strategies in the online Supplementary Materials), they picked six for discussion, which are reproduced in Table 8. Of these, two strategies offer substantial protection (*24* and *25*). A *pre-pandemic* vaccine (efficacy as noted on the previous page) in combination with household prophylaxis and treatment of cases can be effective (strategy *24*). The effectiveness was enhanced by social distancing, which is singly ineffective. A more intensive prophylaxis policy with entry controls can be very successful, and would only require an antiviral stockpile to cover 11% of the population at R_0=1.7 (strategy *25*).

The authors note several caveats about their model and assumptions which are omitted here. Given that the policy case relates to the US (and GB), the objective of the two governments may be to prevent any outbreaks at all. Therefore, the policy of border (entry) controls is of particular significance – a policy included in the best mixed strategy (*25*). This policy has high economic and social costs, and it is reasonable to expect it will be invoked only given carefully considered need. In the modeling for Ferguson *et al*. (2006), this need is not carefully considered. While strategy *25* has an attack rate reduction outcome, this is not associated with pure border control or movement restriction policies *2-4*. Part of the reason lies in the model seeding assumption, via which infectives arrive in the US (and GB). The authors model the epidemic in the Rest of the World (ROW) as a homogeneous SEIR (deterministic model), implying rapid and universal spread. From the incidence of infections per day in the ROW, the authors sample a proportion as the number of 'imported infections per day' into the US, dependent on the intensity and nature of air traffic. Border controls in the Ferguson *et al.* (2006) model arbitrarily begin with some delay and act to reduce the proportion of imported infections by 90-99.9%; eventually compensated for by an increase in world prevalence. The result is a delay in the epidemic peak in the US (pushing of the epidemic curve to the right), but outbreaks do inevitably begin in the US. The onus for limiting them thereafter is on some form of mixed strategy.

In the context of the results above, consider that the objective in the US is *containment*, not just mitigation. In other words, assume it may be possible to eliminate the

possibility of imported outbreaks. Separately, assume that limiting the import of infectives has larger policy benefits beyond delaying a peak. If true, these have implications for policy in countries at risk of secondary outbreaks of pandemic influenza.

For the former possibility, note that international spread may not be instantaneous *or* homogeneous in terms of the risks posed for importing infectives into a country. Some regions may be seen at high risk of pandemic influenza outbreaks (e.g., rural SE Asia). A global influenza transmission model based on air-traffic – as admitted by Ferguson *et al.* (2006, Supplementary Information) – would better predict the risk of importing infectives. A city's risk would derive from the intensity of its connections and links to cities at high risk of outbreaks. Given this heterogeneity in the ROW pandemic transmission, and the learning from SARS, a blanket ban on flights from certain regions, incrementally updated based on new information[49], could indefinitely delay importation of infectives. An airport-based quarantine and antiviral prophylaxis policy (as planned at airports like LAX[50]), could supplement this, targeting any initially asymptomatic travelers detected en route and those they potentially exposed. It can be recalled that of the 29 SARS cases in the US (plus 137 suspect and 19 probable), all imported, there were no reports of secondary transmission[51]. 'Intelligent border controls' may *contain* the US epidemic, i.e., minimize the attack rate.

Second, even if an intelligent border control policy, supplemented by airport-based quarantine/prophylaxis plus movement restrictions only delays the outbreaks, this delay is valuable for pandemic vaccine policies. This is considered explicitly in Germann *et al.* below. The delay may also help healthcare facilities better prepare for the surge in demand that is associated with the peak (if the peak itself cannot be reduced). This is a crucial consideration in actual prevention of severe illness and mortality. In Table 8, 99.9% effective border control (policy *3*) – given a homogenous SEIR transmission model for ROW – obtained a delay of 50 days in the US peak. In the context of a heterogeneous ROW transmission scenario, this delay may be larger and applied to the first instance of secondary cases rather than the peak. Even a 50 day window is valuable for the development and deployment of a pandemic vaccine, for which manufacturers no longer solely rely on egg-based production. As available, this vaccine may be administered to passengers, airport and healthcare workers, and other potential exposures, which may minimize the in-country case attack rate (see policy *18* in Table 8).

4. **Germann *et al.* (2006)**[38]: The United States

Policies compared: The policies compared here are similar to Longini *et al.* and Ferguson *et al.* (2006). The model derives from the former. For an assumed percentage of true index cases (excludes false positives), social targeting of antivirals (TAP variant) is considered at different levels for their preschool, school, and work contacts, as relevant. It is assumed a fixed *60% of cases* and their contacts can always be targeted with antiviral, the constraint being stockpile size. Social distancing is considered – mainly the closure of schools, including preschool and play groups. In addition, movement restrictions are considered. The authors consider a 'dynamic mass vaccination' policy which uses both pre-pandemic and pandemic vaccines as they become available, with distribution schemes either random across the eligible population or prioritizing children. A one-dose policy allows the vaccination of twice as many people, assumed sufficient for achieving efficacy levels for a pre-pandemic vaccine. A two-dose policy confers maximum protection for a well-matched pandemic vaccine. Specifics are given with Table 9 below.

Assumed pharmaceutical efficacy: Oseltamivir efficacy and pre-pandemic vaccine efficacy are as in Longini *et al*. For an assumed *pandemic* vaccine, $VE_S=0.7$ (0.5 for those older than 65), and $VE_I=0.8$, more optimistic than Ferguson *et al.* (2006). Two doses of the pandemic vaccine and one dose for the pre-pandemic vaccine are assumed to be required.

Control objective: Mitigation, as in Ferguson *et al.* (2006), with flavors of containment. In their discussion, the authors suggest policies to slow the spread within the US (delay the epidemic peak) such that mitigation policies can reduce total morbidity.

Table 9: *Effectiveness of policies for control in a US pop. of 281 million* *(Germann et al. 2006)*

	Illnesses : cumulative incidence per 100	
Policy/Mixed strategy	R_0=1.6	R_0=1.9
1. No intervention	32.6	43.5
2. TAP with unlimited stockpile*	0.06 (2.8 mn.)	4.3 (182 mn.)
3. Dynamic vaccination: one dose, random†	0.7	17.7
4. Dynamic vaccination: one dose, child-first†	0.04	2.8
5. Dynamic vaccination: two doses, random†	3.2	33.8
6. Dynamic vaccination: two doses, child-first†	0.9	25.1
7. School closure¶	1.0	29.3
8. Local social distancing¶	25.1	39.2
9. Travel restrictions§	32.8	44.0
10. Local social distancing and travel restrictions¶	19.6	39.3
11. TAP*, school closure**, & social distancing**	0.02 (0.6 mn.)	0.07 (1.6 mn.)
12. Dynamic one dose random vaccination†, social distancing¶, travel restrictions¶§, & school closure**	0.04	0.2

13. TAP*, dynamic one dose random vaccination†, social distancing¶, travel restrictions¶§, & school closure**	0.02 (0.3 mn.)	0.03 (0.7 mn.)
14. Dynamic one dose child-first vaccination†, social distancing¶, travel restrictions¶§, & school closure**	0.02	0.2

* For 60% cases. Policy begins 7 days after pandemic alert (brackets: antiviral supply needed, in millions)
† 10 mn. doses per wk. for 25 wks., timed so individuals develop immunity on the date of first US index case.
¶ Intervention starting 7 days after pandemic alert   ** Intervention starting 14 days after pandemic alert.
§ Reduction in long-distance travel to 10% of normal frequency, occurs during entire simulated epidemic

Discussion: Table 9 above closely follows the results as reported in the study. Given comparable vaccine *efficacy* to Ferguson *et al.* (2006), the *effectiveness* of dynamic vaccination policy is very high, lowering cumulative incidence (analogous to the attack rate in this context) to 0.04% at R_0=1.6. The best vaccination policy (**4** in Table 9) prioritizes children and is antigen saving by using only one dose. The distribution is timed such that a proportion of the population has some immunity before the first US case. This proportion depends on the number of weeks by which the vaccination program precedes a global outbreak. In comparison, Ferguson *et al.* (2006) model pre-pandemic vaccine administered to a fixed 20% of the population while prioritizing children, which in combination with other interventions can lower attack rates to as low as 1% at R_0=1.7 (not shown in Table 8).

The threshold effects of the basic reproductive number R_0 are very intense in their study for the increase from 1.6 to 1.9 (results for R_0>1.9 not shown in Table 9 imply single policies will fail). But even at higher levels of R_0, combination strategies with a dynamic vaccination policy perform very well, and also limit the need for large antiviral stockpiles.

The authors consider the potential feasibility issues for TAP carefully and find that the policy – though successful at lower levels of R_0 – requires many onerous assumptions of policy preparedness and is vulnerable to uncertainty. School closure and other social distancing are considered, but do not generate sufficient benefits given the social cost. However, a TAP policy in combination with social distancing (strategy **11**) is valuable for containing the attack rate, such that a larger proportion of the population remains eligible until the point a well-matched pandemic vaccine becomes available (strategy **13**). Such a mixed strategy reduces the overall caseload dramatically. This combination was as suggested by the critique of Ferguson *et al.* (2006) above. Germann *et al.* do not model border control policies for achieving the delay useful to deploy a pandemic vaccine.

The combination strategies generate benefits in their model due to the nature of the 'dynamic vaccination policy'. Here more efficacious vaccines for larger proportions of the population become viable (both due to an antigen sparing single dose regimen and because

less people are symptomatic due to TAP/social distancing). This process is realistic, as the incrementally available information on the circulating human influenza A strains is currently used to update the seasonal vaccines. In their model, a pre-pandemic vaccine (based on precursor strains) could be available up to 2 months before the global outbreak, and a better matched vaccine up to 2 months later. The details were not provided of the initiation point within this timeframe for the two vaccine types for the dynamic vaccination related policies/strategies in Table 9 (*3-6*, *12-14*). Despite this unknown and connected uncertainties in the efficacy of a pandemic vaccine, or in the dose dependent delay between vaccination and full efficacy, or in the actually feasible dosage to achieve significant population coverage; Table 9 encourages a vaccination using mixed strategy.

Studies of the effectiveness of specific policies

1. **Vaccination policies:** *Distribution and dose size choices*

The discussion of vaccination policies in the comparative studies above introduced two issues in choosing an appropriate policy: type of targeting (distribution) and the size of the vaccine dose. The former issue emerges from a debate on allocation policies for a vaccine (pre-pandemic or pandemic) based on the assumption that antigen production ramps up slowly after a candidate strain is identified. Targeted distribution in this context may prioritize the most infectious individuals (children, e.g., policy **16** and **17** Table 8) to limit total sickness, or those most likely to face serious health complications up to death as a result of sickness (older people, e.g., policy **19** in Table 8). Separately, the issue of dose size also relates to a shortage of antigen, but with a different flavor. This issue has emerged from the realization that even given planned increase in global vaccine production capacity (currently at 350 million doses of trivalent seasonal influenza vaccine[24]), antigen supplies will be lower than required to cover significant proportions of population in countries at risk of initial outbreaks. In this context, studies have compared the effectiveness of a vaccination policy that uses a smaller dose – sacrificing some efficacy for higher coverage of the population – to a policy using the maximum dose for smaller coverage. Both distribution/targeting and dose size issues are considered below.

Choices in targeted distribution: In the US, the National Vaccine Advisory Committee and the Advisory Committee on Immunization Policy (NVAC/ACIP) recommend a scheme that puts highest priority on high-risk individuals 0-64, vaccine and health-care workers (HCW), government leaders, and pregnant women. The next tier of priority applies those

above 65, moderate risk individuals, and infants[52]. Healthy individuals aged 2-64 receive lowest priority. This plan is evaluated by Meltzer *et al.* as discussed further below.

Bansal *et al.* (2006)[22] evaluate four targeted vaccination policies for pandemic influenza: a mortality-based variant that targets infants, adults, and HCW; a morbidity-based variant that targets school-aged children and school staff; a mixed strategy that targets groups with high attack rates (children) and high mortality rates (infants and adults); and a contact-based variant that removes a fraction of the most connected individuals in the modeled social network. The authors use epidemic contact network analysis to model the spread of pandemic influenza, varying the transmissibility of the viral strain by varying ρ (the transmission probability per contact). The model is based on demographic and social network data from Vancouver (Canada). Approximate results based on the graphical depiction in their paper are shown in Table 10 for discussion.

Table 10: Effectiveness of different targeted vaccination policies (Bansal et al. 2006)

Targeting policy	Proportion effectively vaccinated*, % (Fraction of available vaccines given to the group, %)					Attack rate, % (Mortality rate, %) at ρ=0.15§
	Infants	Children	Adults	Elderly	HCW	
No vaccine	-	-	-	-	-	≈62 (≈0.35)
Mortality-based	≈75 (≈25)	-	≈50 (75)	-	≈40	≈50 (≈0.26)
Morbidity-based	-	≈50 (≈95)	≈2 (≈5)	≈2	-	≈38 (≈0.25)
Mixed	≈10 (≈5)	≈25 (≈45)	≈10 (50)	-	≈20	≈40 (≈0.25)
Contact-based	-	≈40 (≈80)	≈2 (≈8)	≈5 (≈10)	≈100 (≈2)	≈20 (≈0.22)

* Product of the implemented coverage level (% of each group's size) and the group-specific vaccine efficacy
§ Transmission probability for human influenza A per contact, assumed linearly related to R_0

The authors model the results of vaccination as full protection, and those 'effectively vaccinated' (see note to Table 10) are removed from the epidemic network. Therefore, the contact-based policy, which is very effective for limiting mortality and morbidity at ρ=0.15, should be disregarded as it is very model specific. The morbidity-based policy is effective on both counts at lower transmissibility. At higher levels of ρ not shown in Table 10, the mortality-based or mixed vaccination policies are more effective to limit mortality.

Overall, this study reiterates the insight that effectiveness depends on the intent of vaccination policy (limiting morbidity or mortality) as well as the base infectiousness of the pandemic viral strain. Morbidity-based policies are in essence transmission-limiting, as children and others in school settings could be most at risk of infection as well as most infectious for reasons previously described. As a result, the policy could lead to herd immunity that unambiguously lowers eventual mortality. The child-prioritizing vaccine policies incorporating some delay in Ferguson *et al.* (policy **16** in Table 8) and Germann *et*

al were effective and robust to infectivity. However, when delays in the start of the policy are introduced in the model in Bansal *et al.*, prioritizing children has higher value (in terms of limiting the attack rate) only if compensated by low to moderate levels of ρ (<0.11). Multiple reintroductions of the pandemic also favor a mortality-based policy in Bansal *et al*. Their results are specific to the particular urban setting (Vancouver), and prone to model and parameter uncertainty (e.g., over the value of ρ). Therefore, the choice of mortality-based vs. children-based policies should be driven by context, as well as early estimates of R_0 in a pandemic. Also, these insights on targeting do not reference feasibility issues that relate to the overall antigen availability. This is considered next.

Choices in vaccination dose size: Riley *et al.* (2007)[24] evaluate the effectiveness of an antigen-conserving vaccination policy that covers a larger share of the population but at a lower level of protection. They use clinical trial data on efficacy for three pre-pandemic vaccines (including one licensed by the US government). The clinical trial data provides a range of dose sizes that produced some immunogenic response, specified from minimum to maximum antigen volume. As an example, for the US government licensed vaccine this range was 7.5-90μg[23]. The model is mathematical, relating the attack rate to the protection offered by a vaccine and to R_0 (for a homogenous population). In their most involved model variant (a 'leaky' policy), they assume that the vaccines could offer partial protection in some individuals even with the maximum dose, i.e., the population is heterogeneous in the efficacy of the vaccine. In this variant, the model is calibrated to a predicted attack rate (73%) at R_0=1.8; derived from the simple, homogenous form of the model[53].

For all three vaccines, increasing population coverage with the minimum dose leads to lower infection attack rates. As an example, assume that for one of the vaccines, an adjuvanted influenza A (H5N1) vaccine, there is a stockpile such that the maximum dose allows coverage of 20 out of 300 million Americans and yields an attack rate of 67.6%. Then the authors find that for a heterogeneous population using the minimum dose is optimum, producing an absolute reduction in the attack rate of 8.9 percentage points. Even if healthcare workers are prioritized (up to 45% reservation of the same stockpile) the minimum dosing policy would still be superior, reducing the attack rate by 4.8 percentage points. The insights are as usual subject to the caution that the clinical trials involved poorly matched pre-pandemic vaccines, and that the dependence on R_0 in the involved model variant is a source of uncertainty. These caveats are discussed in a related review note[54].

2. **Community policies**: *Case isolation & contact tracing, social distancing, quarantines*

Community-level nonpharmaceutical interventions (NPIs) figure across the containment and mitigation focused studies above. These form the core of the WHO *Interim Protocol*, aimed at developing countries, as well as the DHHS/CDC[7] plan, where different NPIs are recommended based on expected severity. Since it is difficult to predict the severity in advance, the prioritization may be difficult to implement. There are two reasons to consider NPIs in further detail. First, community-level policies are likely to be of importance in resource-poor settings. Second, the comparative studies produced a dichotomy that goes beyond containment vs. mitigation focus. Generally, social distancing and quarantines were effective and robust to an increase in R_0 for containment in SE Asia (see policy *f* in Table 6, and compare policy *4* to *1* in Table 7), but not for mitigation in the US (e.g., review policies *7-10* in Table 9 at R_0=1.6, and then across R_0:1.6-1.9). This counters the logic that policies successful at containment should also mitigate successfully. If there are factors beyond methodology (e.g., different objectives or core assumptions) that make community-level policies ineffective in some contexts, it should interest policymakers. This issue can be examined with recent studies that specifically analyze these policies, beginning with analyses of historical data on community-level nonpharmaceutical policies.

Historical evidence: A WHO review[55] surveyed the accounts from the 1918 pandemic (when almost all control policies were nonpharmaceutical) and found public health officials of the period skeptical on the success of quarantines or case isolation. Early case detection and isolation supplemented by movement restrictions had more impact on attack rates in closed settings (e.g., military barracks and college dormitories). Social distancing, especially school closure, may have more value, as suggested by studies of seasonal influenza epidemics in Israel and France[55]. However, in some contexts school closure or holidays can lead to children becoming more mobile in the community, thus leading to higher transmission. The WHO review provides some evidence for this from US cities during the 1918 pandemic. In this case, school closure without additional household restrictions on children may not be as effective. In general, the WHO review concludes that North American and Australian data from the 1918 pandemic indicate that community level nonpharmaceutical measures were not 'demonstrably' useful historically.

Recent studies of NPIs during the 1918 pandemic have examined the evidence. Hatchett *et al.* (2007)[56] analyze data from 17 cities that implemented some form of NPIs and

varied in their peak or overall mortality rates during the fall 1918 pandemic wave. The most common NPI configurations included social distancing and case isolation. The authors found that school, theatre, and church closures were most effective. By statistically analyzing the difference in pandemic outcomes given the policies employed, the authors conclude that timing of the interventions was a key explanatory factor for peak mortality, but not overall mortality. In other words, city governments who initiated interventions sooner after an outbreak began in their jurisdiction were able to 'flatten' the epidemic curve, and also limit the total deaths[57] (a proxy for the infection attack rate). Again, this points to how the demonstration effect of prior outbreaks allows cities with later outbreaks to time their interventions better and achieve overall reductions in *R*.

The reduction in overall mortality could have been much larger if the effective interventions had continued for a longer period. As it happened, cities with effective community interventions protected their citizens from the first wave, but this kept a larger number susceptible. When the interventions were lifted prematurely (usually after mortality rates began to decline), the cities experienced a second wave as infectious individuals restarted an epidemic. In comparison, second waves were relatively smaller in cities with poor policy implementation in the first wave, even if such cities may have had larger overall mortality. Epidemiologically, the cities with effective but prematurely lifted interventions did not reduce *R* to a low enough level to achieve substantial herd immunity. These insights are corroborated by Bootsma *et al.* (2007)[34], who fit the 1918 data to an SEIR model and test similar hypotheses. Bootsma *et al.* conclude that transitory community-level policies have the potential to reduce attack rates (and hence mortality) by 30-40%, but higher reductions require sustaining the social distancing and movement restriction policies for longer durations, imposing greater social and economic costs.

Besides the insights on timing and length of policies, both studies make a point not considered before about individual reactions. The 1918 data indicates that rising mortality rates caused people to react by reducing risky behavior. This manifested most visibly as lower rates of social contact, and possibly as greater personal hygiene or mask-wearing. Such 'prevalence-elastic' behavior has been theorized before for other diseases. In statistical analysis, such reactive behavior plus indirect effects from family members reducing contacts to stay at home to care for the sick boosts the effectiveness of social distancing policies. But, in realistic epidemic settings, the results of over-dependence on reactive behavior may be difficult to predict. Poor information availability – especially on mortality

which is the most observable indicator but occurs with a lag to transmission – may mean that private reactive behavior is too little and too late. Prompt and trustworthy public health communication is required to reinforce the positive factors, as seen during SARS.

Prospective modeling: Carrat *et al.* (2006)[58] using French data on effective contact rates in an influenza epidemic, simulate community transmission of pandemic influenza and find local nonpharmaceutical policies more effective compared to antiviral prophylaxis. A policy of contact tracing and confinement implemented for 70% of reported cases worked well to limit the attack rate, as did general school and workplace closure started after a threshold of infections was reached. These results are in tune with the insights from the historical studies, but confound those from Ferguson *et al.* (2006).

It is likely that during a pandemic substantial number of secondary infections may occur in households, where the index cases acquire the disease in community transmission. Wu *et al.* (2006)[27] compare the effect on the overall attack rate of community-level policies that reduce such within- and between-household transmission. The model used is a 'stochastic SEIR' simulation, which lacks a spatial component unlike the other comparative studies. Wu *et al.* make the claim that rapid spread will make this issue less relevant[47].

Table 11: Effectiveness of different NPIs/mixed strategies in a pop. of 1 mn. (Wu et al. 2006)

Policy/mixed strategy	R_0=1.8, θ=30%	
	Attack rate %	Peak Q*
No intervention	74%	0
Q: *Quarantine* - segregation within their own homes of a complying household contacts of suspected case	49%	9.6%
QI: *Q + Isolation* (*I*, complying symptomatic individuals are removed from their household to a separate facility)	43%	7.1%
QA: *Q + Antivirals* (*A*, complying symptomatic household members take 2 doses of antiviral and symptom-free members take 1)	44%	0.5%**
QIA: *Q + I + A* for all complying individuals, as appropriate	40%	6.2%
QIAC: *Q + I + A + Contact tracing* (*C*, symptomatic and isolated individuals name people they may have infected, and those are notified and asked to take precautionary measures *Q* and *A*)	34%	≈13.5%

* Peak % of the population living in conditions of household quarantine ** 3.9 doses per person in the population

The authors find a significant role in mitigation for NPIs at moderate levels of compliance even without TAP. At similar compliance levels in Ferguson *et al.* (2006), quarantine had lower relative effectiveness (policy **12**, Table 8). Antiviral prophylaxis reduces the number of people in quarantine, thus reducing the social and ethical burden. However, for the baseline simulation (Table 11) at 50% compliance for the policies and R_0=1.8, the antiviral stockpile required is quite large. The proportion of presymptomatic or asymptomatic transmission, θ, was assumed to be 30% in the baseline simulation[60].

The major contribution of their study is highlighting aspects of feasibility for community-level policies. First, they explicitly compare their assumptions to those in the studies by Germann *et al.* and Ferguson *et al.* (2006). While Wu *et al.* assume 67% of cases will show symptoms and be reported (vs. 50% in Ferguson *et al.* and 60% in Germann *et al.*), they conservatively assume a 50% compliance with subsequent policy interventions. The other two studies mostly model higher levels of implementation with their reported cases. Wu *et al.* rule out contact tracing as infeasible in large networks (even if effective as **QIAC**), though Germann *et al.* allow that TAP for 60-100% of contacts outside the household will be feasible in early stages of a US epidemic. Second, Wu *et al.* consider compliance more carefully. For policies that confer immediate benefits to the individual (**QA**, as well as **QI** since guaranteed access to antiviral is assumed in case isolation) compliance may be higher, while lower for quarantine policy. Compliance with isolation would be affected by the characteristics of the index case and household structure. Third, with the peak quarantine rate as an outcome, the authors discuss feasibility in terms of the demands various policies place on authorities to ensure essential supplies for the restricted population. Finally, they point out that strategies involving a case isolation policy assume too easily that facilities to effectively isolate individuals will be available, especially when urban transmission occurs on a significant scale. If antivirals are available and useful, they may be considered in lieu of isolation (i.e., **QA** vs. **QI**). Where antiviral stockpiles are limited, the use of large-scale case isolation may be needed – a contingency to be planned and prepared for.

Discussion: The results above indicate NPIs have a role in *mitigation* at moderate R_0 and θ, even without TAP or vaccination in tandem. The timing requirements are not more onerous than for drug-based strategies, but the duration NPIs have to continue to minimize the GAR – based on historical experience – may have significant societal costs. The dichotomy in NPI effectiveness seen in the modeling studies across SE Asia and the US may be due to model-related factors. The early use of NPI intervention, limited seeding, and the lack of reintroductions in the SE Asian models may explain the observed effectiveness of NPI in containment. In contrast, given some delay in intervention start, multiple infection seeds, reintroduction risk, and widespread transmission as in the US models, a persistent decline in susceptibility and infectiousness through vaccination and/or TAP is preferred (as *11* in Table 9). However, pharmaceutical policies are resource and capacity intensive, and may be overkill in conditions of moderate R_0 if a strategy of cooperative international air travel restriction mixed with border control/airport quarantine is probable. In this case, for

countries at secondary risk, the policy objective resembles early containment. This implies that the choice of stringent, layered NPIs at airports and *reactive* NPIs in communities (as *15* in Table 8) could minimize the attack rate. This strategy has *a priori* lower social cost if a smaller epidemic size is assumed since widespread school and workplace closure are avoided and unfeasible contact tracing during sustained urban transmission is not required. However, in situations of higher infectivity (R_0), the primary benefit of NPIs will be to delay peak transmission such that a well-matched vaccination policy becomes viable. However, vaccination contains the assumption of adequate antigen stocks in resource-poor settings.

3. **International air travel restrictions and advisories**

The reduction in the volume of and entry-points for incoming infectives would eliminate multiple outbreaks seeding; increasing the probability of waging containment battles at airports and high risk communities rather than a large-scale mitigation war. This could be achieved with entry and travel controls, which could also delay an epidemic peak in a country. Depending on the delay, this would enhance the viability and timeliness of other control policies from NPI to vaccination. For pandemic influenza – given higher θ than SARS – symptom-based entry screening is not expected to be an effective preventive measure[18]. Instead, the cancellation of flights to and from certain (or all) cities worldwide could be considered by each airport. Retrospective modeling of the 1968-69 pandemic[31] suggested a role of the international air network in the spread of influenza. Recent work by Brownstein *et al.* (2006)[61] analyzed the 'natural experiment' of the post 9-11 shutdown in US air traffic, and found it connected with a delay in the seasonal influenza peak in 2002.

However, control in an actual pandemic situation with higher transmissibility and susceptibility compared to interpandemic influenza needs to be evaluated. The success of international air travel restrictions (IATR) for pandemic control depends on a proactive WHO role, the nature of the initial outbreaks, the identification of highly connected cities that may exchange infectives (which may be isolated from each other), further analysis of the SARS experience, and some voluntary self-restriction by travelers who expect to have been exposed to infection. There is considerable uncertainty if IATR will be effective. Modeling studies have attempted to bring greater predictability to the issue.

The comparative modeling studies did not model IATR or entry controls in great detail. Full border controls – 99.9% effective in reducing imported infectives – in Ferguson *et al.* (2006) only delay the epidemic peak. As noted previously, if most countries cooperate

on air travel restrictions and information on outbreaks is widely and immediately shared, the global circulation of infectives may be lower. In other words, the global prevalence of pandemic influenza need not be ever rising to diminish the effect of border controls. Germann *et al.* make a simpler assumption that while international air travel will be the dominant mode of infection seeding in most countries, achieving 'impenetrable' borders will be prohibitively expensive. However, such considerations did not prevent studies from evaluating NPIs or drug policies that affect many persons or require significant resources.

Two recent studies[62,63] model international air travel within a highly connected or nodal set (105-155) of major global cities using real flight data and demographics, and attempt to simulate the global spread of pandemic influenza. They introduce similar IATR policies in similar model settings, but obtain differing results. Both models use a stochastic SEIR model for the epidemic in each city, with the proportions of people in various states in a city affected by the number of travelers entering and departing each city. The proportion of the air passengers in the various disease states traveling between two connected cities is identical to the proportion of people in the states in the city of origin. Sequential travel bans – considered more realistic than a simultaneous global shutdown of the air network – begin in each city after a threshold of cases is recorded. This threshold is higher for the first city to have a major outbreak. This framework is biased such that epidemics in each city in the modeled network are near inevitable (containment is not an objective), but the peak in each city may be delayed by the sequential flight bans. This effect manifests in the city imposing restrictions, but also in cities connected to it yet to reach their intervention threshold.

Cooper *et al.* (2006)[62] set the threshold at 1,000 symptomatic cases for the originating city, and 100 for each city thereafter. Even when 99.9% of air traffic was suspended, there was no containment (loosely defined as the low probability of an epidemic in at least some cities), and delays in peaks were small and insignificant for the viability of other control policies. This result was robust to assumptions about the originating city in their model, the timing of the first outbreak, the initial susceptibility of populations, and infectiousness (only at very low R_0 values was there significant delay).

Epstein *et al.* (2007)[63] use a flat threshold of 1,000 cases for their baseline modeling of sequential flight bans. The results obtained are slightly less bleak on the effectiveness of IATR – a 95% reduction in flight volume would produce a delay of two to three weeks at R_0 =1.7, and hence a reduction in the total global cases after 6 months. At the baseline, there is

little or a worsening effect with IATR, depending on the timing of the first outbreak[64], on total cases over the entire pandemic. The results are sensitive to the originating city (delays are larger with Hong Kong than London). The authors model the use of a vaccination policy in tandem with IATR and find that the mixed strategy greatly boosts the effect on the attack rate. Adding a policy of 0.1% daily vaccination – which can be applied to larger number of people thanks to the delay in epidemic peaks achieved by IATR – would reduce the number of cases in urban areas of the US from 102 million to 57 million. This insight follows discussion previously in the context of the studies by Ferguson *et al.* (2006) and Germann *et al.* The 2-3 weeks delay in itself would also allow many more cities to get ready with NPIs.

These results are not encouraging for IATR proponents looking for containment of a global pandemic, i.e., major reduction in the overall number of cases. It is uncertain whether 95% reduction in travel volumes could be achieved with sequential restrictions, but this would at least generate delays sufficient for other policies. Epstein *et al.* have already considered that the threshold for intervention would reduce, as cities further down the timeline react faster. This factor did not change their conclusions in a major way. But, future research with stochastic SEIR/air travel models could address some lacunae.

First, the two studies assume that in the no-intervention scenario the air travel network remains stable in flight and passenger volume until sequential IATR begins, which implies significant circulation of infectives and the spread of pandemic influenza. However, flights cancellations may begin earlier in cities other than the originating city, before their 100/1000 cases are recorded. This is because flight cancellations at one airport affect airline operations, which means disruptions can spread to other airports, especially major hubs[65]. This effect may bring down global *flight volume* much earlier than predicted by the IATR timeline. When flight bans are imposed, they enhance the airline-related ripple effect. The amount of early reductions in flight volume will depend on the location of the first outbreaks and the associated airlines. Second, effects deriving from individual volition can bring down *passenger volume*, e.g., decisions not to fly at all based on assumed risk of infection on flights or at airports. Due to such prevalence-elastic behavior, many trips except for repatriating visitors may be cancelled, especially in countries near an outbreak zone. This would boost the effectiveness of sequential restrictions in a similar way.

Third, further research is required on the benefit of passenger- and airport-based interventions that are dependent on a global information criterion and begin much earlier

rather than indicated by a case-rate threshold in the city an airport serves. The decision triggers in the US pandemic plan are an example of such early policy activation[7]. The effective contact rate of passengers may be easily modified with in-flight risk communication, airport-based interventions (issue of PPE, antivirals), etc., that could begin very soon after the first outbreak is reported anywhere in the world. Though asymptomatic and symptomatic transmission may still occur on a flight, and later with community contacts (in the absence of passenger quarantines), the compartmentalization of passengers – in the language of uniform mixing SEIR models – may mean R_0 should decline in cities further down a timeline that benefit from more information. Such passenger- or airport-focused interventions that follow the principles of 'targeted layered containment' would improve with better global outbreak surveillance, WHO coordination, and airline participation. By reducing travel volume, IATR would boost the effectiveness of these policies. As a result, the chances would be higher that more cities contain their epidemics and that the delays in peaks in others are longer.

4. **Therapeutic use of antivirals with global redistribution of stockpiles**

Colizza et al. (2007)[80] investigate a global mitigation strategy for pandemic influenza involving the therapeutic (rather than preventive or prophylactic) use of antivirals. Their international model links 3,100 urban areas across 220 countries in an airline network, and they use a stochastic SEIR model to simulate the epidemic within each of those urban areas. Effects due to seasonality in influenza epidemics are factored in for parameters on infectiousness. Based on the results in other studies[62,81] they discount the possibility of IATR obtaining significant delays in pandemic peaks or reductions in overall attack rates. As a result, they focus their modeling of interventions on antivirals.

Assumed pharmaceutical efficacy: Neuraminidase inhibitor antiviral efficacy for reducing infectiousness (AVE$_I$) is 0.62, as in Longini *et al*. Further, Colizza *et al.* assume that the average length of the infectious period for treated and infectious individuals is reduced by one day. Given the proportion p_{AV} of symptomatic individuals who can be feasibly identified and treated per day, the overall reduction in the infectiousness of treated individuals ranges from 30-50%. With baseline values, they approximate the following relation: $R_{av}=R_0/1.3$, where R_{av} obtains based on the unrealistic expectation of timely and widespread therapeutic use of neuraminidase inhibitors from the start of the pandemic.

This reduction is lower than in $R_{av}=R_0/3.6$, proposed for the prophylactic use of antivirals of similar base efficacy[27].

Policies: The authors consider two scenarios for an antiviral policy. In the first 'maximal coverage' scenario, every urban area has sufficient antiviral stockpiles to treat cases according the prevailing treatment protocol p_{AV}. The second, realistic 'limited supplies' scenario has two flavors. The *uncooperative* flavor of this scenario has a limited number of rich ('prepared') countries with stockpiles sufficient to treat 10% of their own populations with antivirals, and two other frontline countries – Vietnam and Thailand for this study – receive some stockpiles as well (up to 10% of the population). In the *cooperative* flavor of the limited supplies scenario, the prepared countries donate from one-tenth (cooperative strategy I) to one-fifth (cooperative strategy II) of their antivirals into a global stockpile for international use as needed. In this cooperative world, symptomatic cases receive treatment as per the p_{AV} as long as the global or country stockpile has drugs. Results for effectiveness in containment (Table 12) and mitigation (Table 13) are below.

Table 12: Probability of a global pandemic one year after seeding in Hanoi in October

Antiviral policy scenario (p_{AV} = 50%)	R_0 = 1.5		R_0 = 1.9		R_0 = 2.3	
	No outbreak	Outbreaks in >100 countries*	No outbreak	Outbreaks in >100 countries*	No outbreak	Outbreaks in >100 countries*
Maximal coverage	≈66%	0%	≈40%	≈60%	≈30%	≈70%
Limited coverage, uncooperative	≈66%	**≈16%**	≈40%	≈60%	≈30%	≈70%
Limited coverage, cooperative –I, II	≈66%	0%	≈40%	≈60%	≈30%	≈70%

* Probabilities not shown for global outbreaks in the ranges: 2-10, 11-50, 51-100 countries.

Table 13: Average cases per 1000 globally after seeding in Hanoi in October (Colizza *et al.* 2007)

Antiviral policy scenario (p_{AV} = 50%)	R_0 = 1.5	R_0 = 1.9	R_0 = 2.3
Maximal coverage	0.01	35	184
Limited coverage, uncooperative	16	222	397
Limited coverage, cooperative -I	0.01	126	305
Limited coverage, cooperative –II	0.01	97	287

Discussion: Containment success is poor across the distribution scenarios for any R_0 greater than 1.5. If sufficient stockpiles of antivirals are available (2-6% of the global population) and a cooperative strategy is maintained with timely redistribution of the stockpiles of prepared countries, then mitigation even at a R_0 of 1.9 is possible. Based on details not shown in Tables 12-13, the cooperative system results in delays of the peak of the global pandemic for as much as a year for R_0 between 1.5-1.9, which is sufficient for the development of a well-matched pandemic vaccine. The amount of redistribution of antiviral

stocks involved is modest, reducing the availability of stockpiles in the prepared countries to 8-9% of the population from the prior 10%.

Cost and cost-effectiveness of policies

Direct costs: The direct cost of vaccine and antigen production/procurement and distribution for the overall health sector could be estimated for different stockpile and coverage scenarios, though rigorous studies of these costs are yet to be initiated. The affordability of large-scale vaccination and antiviral prophylaxis policies remains out of the reach of many developing nations. These costs and potential financing mechanisms require urgent research given the conclusions from the studies reviewed so far.

For NPIs, it is unclear what the direct costs should include: social, economic, and resource costs will all be incurred. For the purposes of this review, the economic costs of NPIs due to reduced demand or due to workforce disruptions (absenteeism and productivity loss due to telecommuting where this occurs) will be considered *indirect* costs. Direct costs to the local and central public health systems of enforcing case isolation, contact tracing, movement restrictions and social distancing may be very large if these continue for long and occur in large cities. As noted by Wu *et al.,* when a quarantine policy is selectively enforced for the households of cases and their contacts, the local authorities may have to ensure a supply of essential supplies for the quarantine to be ethical. Also, the disruption to the normal routine of people has a social cost which is difficult to compute. Such cost issues are a prime factor leading the DHHS/CDC plan to be circumspect about the recommended NPIs at different forecast levels of severity[66]. NPIs would be instituted in an actual pandemic regardless of such recommendations. The DHHS/CDC plan institutes centralized coordination which can reduce social costs. However, severity is difficult to predict in a way meaningful to public health plan activation, and the cost of any mistakes in terms of an inefficiently contained/mitigated epidemic with higher illnesses and deaths would be worse. Hence, the results of the modeling studies reviewed so far on the effectiveness of policies at different levels of transmissibility and feasibility are appropriate starting points for taking decisions based on the local context.

Indirect costs: The indirect costs emerge from disruptions of domestic economic activity from the NPIs in the intervention scenario, as well as the reduction in international travel and business due to IATR. The costs of IATR were estimated by Epstein *et al.* for the

US *and* found to be lower than previously assumed. The per annum cost of major IATR is in the vicinity of 0.8% of GNP, vs. a potential reduction in the overall caseload reduction of about 45 million cases in urban areas in their model. Whether this loss in GDP is low from a policy standpoint is a cost-effectiveness question. Such questions are controversial where it answers would require analysts to value the mitigated morbidity and mortality and the long-term economic benefits therein. The latter estimates are available from other studies[8] which have been reviewed extensively elsewhere. Costs of NPI related reduction in demand in a city with social distancing or quarantines have been estimated as well in prospective economic modeling. The experience from SARS, where there was a quarterly reduction in the demand for services and in travel/hospitality industries in the affected cities (Hong Kong, Taipei, Beijing) are the basis for such estimates. Some of the estimates from outbreaks in locations where social distancing and/or quarantines were instituted are reported in Table 14.

Table 14: Disease outbreaks: estimates of business losses due to local disruption

Location (Year)	Length of disruption	Sectors Affected	Estimated Loss	Disease
Surat, India (1994)	Two-three weeks	Local demand (during the festival season)	US$ 260 mn.[67]	Pneumonic plague (1391 cases[*], 52-68 deaths)
Beijing, PRC (2003)	Nine months (April-Dec)	Decline in domestic tourists (87%) and cancellation of Labor Day holiday (13%)	US$ 2.8 bn.[68] (7.5% of 2002 Beijing GDP)	SARS (2521 cases, 190 deaths)
Taiwan (2003)	Two months (May-June)	Tourism revenue (domestic and international)	US $350 mn.[69]	SARS (346 cases, 37 deaths)
Hong Kong, PRC (2003)	Three months (March-May)	Local demand (metric: total retail sales)	-6.1% in March, -15.1% in April, & -11.1% in May, year on year[70]	SARS (1755 cases, 398 deaths)

[*] Probable cases for Gujarat state.

While the effects on the local economy dissipate in subsequent financial periods as demand picks up and workers report back, in the shortest term the possibility of severe losses to the local business sector can 'stay the hand' of city governments, imposing unnecessary added risk of epidemic escalation[71]. This is crucial for effectiveness, since studies have repeatedly stressed timeliness and completeness for all control policies.

Cost-effectiveness studies: Meltzer *et al*. (1999)[9] estimated the net returns to vaccination policies in the US – targeted (mortality-based) or mass vaccination – using an assumed range for cost of vaccine per person. Vaccine efficacy was defined in terms of reductions in the health outcomes (death, types of healthcare utilization) in the vaccinated group. A high scenario of vaccine efficacy reduced death in those vaccinated by 60-75% (higher in the younger age groups) from baseline, and hospitalization from 50-55%.

Benefits of vaccination were estimated as the avoided costs of excess mortality, accounting for 83% of economic losses in the no-intervention scenario, as well as the avoided costs of excess healthcare utilization (17% of losses). This approach did not include avoided costs due to demand or supply disruptions discussed above. The former costs were estimated using the present value of lost earnings, varied by age group, while the costs of hospitalization varied depending on the severity of symptoms, and were aggregated from the numbers of individuals with different health outcomes from the baseline scenario. The following discussion should be read in light of the review of their 'mathematical scenario' methodology in a previous section.

If maximizing calculable net returns is an objective (other objectives: minimizing the total case attack rate, minimizing risk for death) then the following results apply. Since the largest amount of losses in their model emerged from deaths, a policy aiming to reduce total cases/deaths would produce the largest returns – e.g. a policy aiming at 60% coverage of the US population (320 million doses of vaccine to be produced and delivered in a 2-3 month period). Given resource constraints for such coverage, especially if a well-matched vaccine requires two doses, certain targeting policies are considered. Vaccinating those at high risk of complications – e.g., comorbidities – in the *0-64* age group (a priority under the new NVAC/ACIP plan) would generate higher returns regardless of the cost of vaccine compared to vaccinating those 65+ at *any* risk (also priority under NVAC/ACIP), or those 0-64 not at high risk (lowest priority under NVAC/ACIP).

The results in van Genugten *et al.* (2003)[42] are presented at a higher level of generalization. They find no relative benefits in terms of deaths and hospitalizations of targeted vaccination (for those 65+ and healthcare workers) compared to mass vaccination. Given the higher resource need of mass vaccination, this argues implicitly for targeted vaccination, though the prioritized target group differs from Meltzer *et al.* Results in both studies are highly sensitive to the assumed variation across age groups in attack rates and the proportion at risk for complications. It is uncertain what patterns and risks would obtain in a real pandemic. The challenge of identifying the high risk proportion in the absence of relevant health records in some countries (or even the population distribution of comorbidities) will complicate wider replication of the analysis in the Meltzer *et al.* study.

Ethical dimensions of pandemic control policies

Ethical dimensions of pharmaceutical policies: When drug stocks are limited and a particular control objective is identified, a decision usually follows to target some particular group: contacts of cases, healthcare workers, older individuals, or those most infectious. These decisions are made based on the local context and using both prospective modeling of efficacy and the actual feasibility. As a result of such targeting, some individuals are invariably left at higher risk of illness or death. In healthcare settings, triage is not uncommon, but the ethical implications are complex. In the resolution of these issues, several operative principles are in circulation. So far in this review, the following principles have appeared in the context of drug-based policies:

- 'Save the most lives' – a policy that reduces the GAR and hence the total deaths
- 'Save those most at risk' – a policy that targets those most likeliest to die or be severely sick, without necessarily considering the impact on the GAR
- Maximize the protection of productive life years – implicit when benefits of avoided mortalities are calculated using the present value of lost future income. This calculation usually benefits those younger with more years of work left.

For a particular drug-distribution policy, the connection to the principles above depends on the arguments made. As an example, some of the NVAC/ACIP vaccine priorities – vaccine workers and HCW– could be construed as 'save the most lives' as they aim to keep vaccine production lines and healthcare centers open, eventually raising total mitigation. Other views on the same priorities resemble 'saves those most at risk' (HCW again, plus people with two or more risk conditions, pregnant women, and older persons). After reviewing the NVAC/ACIP priorities, Emanuel *et al.* (2006)[72] propose a 'life-cycle principle' which prioritizes younger individuals over older cohorts, as they have more life stages to pass through. Vaccine and HCW workers are still the top tier of the priorities. An investment variant of the life-cycle principle prioritizes young adults and those in college over children, as they have made more developed life plans. The intent is to save the most 'years of life stages' across the whole population.

A recent WHO consultation on ethical issues raised concerns that prioritizing HCW in a situation of a *severe* pandemic – where it could be assumed that their role has been limited compared to household members – would lead to requests for prioritization from

other professions who also consider themselves indispensable[73]. In a moderate pandemic, the WHO consultation recommendations for vaccination resemble NVAC/ACIP.

Though some age-based prioritization may be considered unavoidable due to projected antigen shortages for vaccination policies, further research is needed on the priorities that match different social objectives of vaccination (and antivirals). Meltzer *et al.* provided an early example of such research. This is required since the effectiveness of targeted distribution policies remain prone to assumptions about the transmissibility of the pandemic strain. For example, in Bansal *et al.* for the objective of limiting the GAR, i.e., 'save most lives', the preference shifts when transmissibility and delays were considered, from a morbidity-based policy that prioritized children to one that was mortality-based and prioritized adults. It may well be ethical to base priorities closely on the available information on age-based mortalities of candidate strains, as well as further modeling of potential effects of transmissibility on vaccination policies.

Ethical dimensions of surveillance and NPIs: As already discussed, isolating cases and quarantining households, and potentially social contacts via tracing, poses profound issues of law and ethics. If the policy is not well explained, voluntary cooperation may be ill-informed and less ethical. The treatment of those in quarantine or isolation must be humane and meet minimum standards of hygiene, comfort and material availability, albeit with due considerations for the outbreak as an emergency situation. For social distancing, quarantine and other NPIs to be ethical and legally sanctioned, they must be evidence-based, documented, and well communicated to the public. Especially for isolation and quarantine, the appropriate legal authority of those declaring and implementing the policy (at separate levels) should be established in all countries[74]. Even where legal provisions for quarantine under medical emergency are well developed, as in the US, a review has found lacunae[74]. This indicates that the issue may require attention in other countries as well. Minor incidents of social unrest witnessed in China during the SARS controls[75] would have been ameliorated with better communication and trust between authorities and the population. Any policy which is unethical at the individual level and/or inequitable in the risks imposed on different groups in the population cannot be called effective.

Ethical dimensions of international pandemic control: The border entry controls and IATR in the situation of worldwide pandemic influenza outbreaks are the rights of individual nations, as well as a potential recommendation from the WHO. However, given

that free movement of people and goods is also a basic right, over-zealous or excessively long duration restrictions against movements from a particular country are challenges to the system. Preemptive travel and import restrictions based on disease risk impose a cost on the targeted country, and while warranted in the situation of pandemic influenza, they require careful consideration. These issues have been tackled in a recent paper[76].

Ethical dimensions of international cooperation: The current system of international cooperation in public health – between governments, private firms, NGOs, the UN agencies and multilateral institutions – functions under the aegis of the WHO and treats prevention and mitigation of diseases of international concern as a global public good. Recently, government and WHO interests in a portion of the system seem to have diverged to the detriment of the effectiveness of future pandemic control. This divergence has ethical dimensions rooted in the sharing of responsibility in and benefits from cooperation.

Alongside early revelation of an outbreak, the biological information on communicable disease strains is critical for the WHO crafting an adequate international response. Under the existing plan for free sharing of disease-related samples (such as seasonal influenza), the WHO has urged countries to make these available to it directly or to its collaborating centers, such that the causative agent can be analyzed and vaccine manufacturers can focus their work. Particularly for the H5N1 avian influenza virus, the WHO established the H5 Reference Laboratory Network, and countries – even if they have the local laboratory capacity to genetically sequence the virus and provide the sequence information electronically – have been urged to contribute human and avian samples.

The WHO views the free sharing of samples as a 'collective responsibility'[77]. The recently released best practices[78], by asking for the open posting by collaborating centers of all the genetic sequence data on the viruses, has given such information the status of an international public good. However, some governments have shown in their actions that they view the viral samples (or both the samples and related sequencing information) as 'national' public goods, which are to be protected from use by other governments or foreign commercial concerns without a defined contract. There are three factors that are driving this dissonance. First, the divergence of the H5N1 viruses into genetically and antigenically distinct clades – like the Indonesian strain – has given specificity to the viral information that is valuable for vaccine development. Additionally, governments – through both national

and international law – have sovereignty over the viral information collected under their jurisdiction, which means that both the WHO and other foreign parties can be excluded[79].

Second, the governments in developing countries have had some justification in questioning the free sharing of sample and sequence information under an 'international public good' system, given the current lack of an accepted global vaccine plan, backed by adequate financing and logistical details. It is a fact that the preponderance of vaccine manufacturing capacity is in the industrialized world, and national pandemic plans ask for securing stockpiles to cover domestic populations first. In such a situation, even if governments honor their international collective responsibility and share information, they cannot be assured that in the use of the information, the information has characteristics of being 'non-rival' in consumption, i.e., affordable global availability of adequate vaccine doses to cover at-risk populations. In fact, global manufacturing capacity for such vaccines, and the technologies involved, remain in the private sector, which would not undertake the massive investments in producing capacity without a compelling commercial logic.

Third, in nationalizing the information, governments stand to make large revenues when they sign privileged commercial contracts. Most developing countries do not have the production capacities or technology to capitalize on the specificity of the viral strain. However, in exclusive revenue-sharing contracts with foreign pharmaceutical companies, both funds as well as technical know-how can be procured, which benefits domestic health security. In contrast, there is a government's net-loss view of the world where other commercial concerns will sell the country a vaccine based on the local strain.

The WHO has been sympathetic to the country governments' concerns, while insisting on the continuing free transfer of sample and genetic sequence information. However, if governments proceed to view viral information as a national good, this may create a major challenge for the response against a potential pandemic of influenza. Some governments have indicated they would honor the global sharing of information if WHO collaborating centers and foreign countries sign Material Transfer Agreements (MTA) that would prevent release of the data for commercial use. However, given that vaccine production capacity remains privatized; such MTAs would be tantamount to crippling the creation of a multi-sourced and robust global influenza vaccine strategy.

Three policy efforts would secure a system of ethical and cooperative international sharing of disease information. First, further urgency is needed for the development of candidate

H5N1 vaccines with true cross-protection across the strains currently circulating. Second, an internationally brokered and financially comprehensive plan is needed that sets agreed-upon pricing and logistical details for adequate doses of a well-matched pandemic vaccine – whatever the national origin of the base strain – for all of the populations at risk. Three, as a part of the international plan for vaccine pricing and logistics, countries that provided the majority of the sequencing and sample data for the candidate vaccine strain could be recompensed with a side-payment of royalties from doses sold in the industrialized world. These potential models of cooperative efforts require urgent research into their feasibility and implementation such that the world is better prepared for a pandemic of influenza.

Conclusions

The studies reviewed in this paper indicate that with adequate preparedness planning and execution it is possible to contain pandemic influenza outbreaks where they occur for viral strains of moderate infectiousness. For viral strains of higher infectiousness, containment may be difficult, but it may be possible to mitigate the effects of the spread of pandemic influenza within a country and/or internationally with a combination of policies suited to the origins and nature of the initial outbreak. Unfortunately, it is very difficult to know in advance how infectious the pandemic-capable viral strain will be; i.e. the basic reproductive number can only be estimated with a lag.

Public health authorities at the start of a pandemic influenza outbreak will attempt to prevent its spread (avoid an epidemic) and will try to isolate and treat the index cases. This is generally referred to as containment. When containment has failed, authorities at the local and national level, even in uninfected countries, will take steps that would prevent rapid and uncontrollable spread, reduce the morbidity and mortality rate of an unavoidable epidemic, and suppress the peak daily rate of cases. Both containment and mitigation policies reduce the overall attack rate, and reduce and delay the peak rate of cases per day. However, by definition, containment begins earlier than mitigation on the time curve of an epidemic and hence has a larger effect on the overall attack rate.

The delay in the peak case rate obtains time for a coordinated international effort to isolate the pandemic strain for vaccine production purposes. However, even if creating a well-matched vaccine is the cornerstone of lasting protection against the particular pandemic-capable strain of influenza, global funding and distribution plans for such a

vaccine remain poorly defined at this point. Stamping out the outbreak in local communities via containment operations, and strictly monitoring and preventing the international communication of the disease is the most preferred course of action[1].

Given the emphasis on timely and efficient containment operations, resource-poor countries at a risk of initial outbreaks may have to invest in containment measures that can address a wide range of scenarios of infectiousness. The need for resources towards such preparedness – keeping in mind the opportunity cost of national and multilateral funds – would lessen if planning focuses on control policies of proven effectiveness. Even if true effectiveness can only be estimated in the field in an actual epidemic due to the unpredictable behavior of individuals and social entities, it is possible to model robust strategies (mix of control policies) that have containment power over a greater range of outbreak scenarios. The more robust is the strategy to pandemic influenza infectiousness, the lower is strain on government capacity, and the greater the possibility of containment at source. At the margin, the probability of containment in resource-poor countries grows if there is a global redistribution – especially of pharmaceutical stockpiles – that matches resources to the risk of preliminary outbreaks. Currently, resources are concentrated in countries at risk of secondary outbreaks, deriving from the focus in such countries on mitigation – rather than containment of initial outbreaks – of pandemic influenza in their society during an 'inevitable' global pandemic.

The modeling studies reviewed in this study indicate a variety of effective control policies and strategies, as well as their variants suited to particular feasibility, cost and ethical standards. These results (summarized below) are a first cut at winnowing the field of potential interventions for containment and mitigation. They are cross-referenced to the known science on theoretical or laboratory efficacy of particular policies. These results indicate the likelihood of containment success in 'frontline risk' countries, given specific resource availability and level of infectiousness; as well as mitigation success in 'secondary' risk countries, given the assumption of inevitable international transmission through air travel networks. However, from the analysis of the modeling results on interventions in the U.S. and U.K. after a global pandemic starts, as well as the results from modeled containment strategies in S.E. Asia, there is a basis for arguing that the emphasis in the secondary risk countries could shift from mitigation towards containment. This follows since a mitigation-focused strategy in developed countries presupposes that initial outbreak containment in these countries will necessarily fail. This is paradoxical if containment success at similar

infectiousness of the virus is likely in developing countries with lower public health resources, based on results using similar modeling methodologies.

Such a shift in emphasis could have major implications for global risk management for diseases of international concern such as pandemic influenza or a SARS-like disease. For example, if international air travel reduction (as a function of globally coordinated policy as well as due to self-restraint of travelers) combined with airport/border control has a strong chance of reducing the entry of infectives into the U.S., then the minimum required size of the U.S. stockpile of antivirals and pre-pandemic vaccines would reduce. This is a reflection of the shift in emphasis from community-level *mitigation* across urban areas towards specific, risk-based *containment* at the level of airport/border areas as well as the traced contacts of international travelers. Given the reduction in the minimum stockpile size in the U.S., the balance could be efficiently and quickly redistributed. This could contain the spread in frontline countries (e.g., in S.E. Asia) for a greater range of infectiousness, reducing the overall global risk of the current wave of the pandemic as well as reducing the possibility of large subsequent outbreaks. This shift in emphasis could reduce the overall global burden of deaths and illness compared to a scenario where the resource-rich countries retain the preponderance of resources in a strategy of 'hope for the best, prepare for the worst'.

Summary of modeling results for containment of outbreaks in frontline countries

Nonpharmaceutical interventions (NPIs) such as quarantine of infected individuals, social distancing measures including the closure of schools and offices, etc., are effective in containing the overall size of an outbreak within a country such as Thailand for viral strains of moderate infectiousness (R_0 less than 1.4)[3]. When NPIs are combined with a policy of targeted antiviral prophylaxis (TAP) of a proportion of the index cases and their contacts, containment can be achieved for viral strains with R_0 as high as 1.7[3]. The policies of TAP or pre-vaccination (with a poorly matched vaccine) are resource intensive given the available laboratory estimates of efficacy of the associated pharmaceuticals. Prophylaxis of a certain number of the population within a ring of a certain radius around an index case or outbreak (drug-sparing geographical TAP: GTAP) can achieve a containment probability of 90% when combined with NPIs, for R_0 as high as 1.9[5]. In a modeled Thai population of 85 million, a feasible strategy combining GTAP using a maximum stockpile of 3 million doses of antivirals mixed with quarantine policies achieves containment for R_0 as high as 1.7[5]. The required stockpile of pharmaceuticals, the effectiveness of policies, and the feasibility of containment

are all sensitive to infectiousness as proxied by R_0. In a model[5], the seeding of the initial outbreak in a sparsely populated rural area can improve the estimated effectiveness of interventions, and reduce cost in terms of resources and economic disruption (e.g., of NPIs).

Summary of modeling results for international containment (travel networks)

Are restrictions in the free movement of international air travel passengers after an outbreak likely to reduce the possibility of a large global pandemic of influenza and delay the peak of a country epidemic? If the policy of sequential travel bans is considered – more realistic than the simultaneous global shutdown of the air travel network – then the results of modeling studies are mixed. In one study (Cooper *et al.* 2006) evaluating a sequential travel ban policy triggered by 1,000 symptomatic cases in the originating city and 100 cases in each city thereafter, the global pandemic was not contained even when 99.9% of air traffic was suspended[62]. This result was robust to assumptions about the originating city, the timing of the outbreak, initial susceptibility and infectiousness (only for very low R_0 was there a delay in the peak of the epidemic). However, another study (Epstein *et al.* 2007), using a flat threshold of 1,000 cases per city to trigger sequential air travel bans found that a 95% reduction in travel volume would produce a delay of two to three weeks in the country epidemics at R_0=1.7[63]. The total global cases would be reduced over the course of a six-month global pandemic timeline. The results were sensitive to the assumption of the originating city. The latter study also found that adding a policy of daily vaccination of 0.1% of the population would reduce the number of cases, e.g., from 102 million at the baseline in the U.S. to 57 million with the air travel restrictions and vaccination. The two to three week delay would also allow more communities and local governments prepare NPIs.

Summary of modeling results for mitigation in secondary risk countries

Suppose the spread of pandemic influenza is inevitable to countries at secondary risk such as the U.S. and the U.K.: i.e., international air travel restrictions (IATRs) do not contain the pandemic, and community transmission in cities of the secondary-risk countries is also inevitable (which assumes airport-based quarantines and controls, and passenger contact tracing as seen for SARS fail, presumably due to reasons such as asymptomatic transmission). In this situation, which policies or combinations of policies (strategies) will reduce the overall attack rate, and reduce and delay the peak rate of cases per day?

One modeling study (Ferguson *et al.* 2006)[37] finds that U.S. strategies combining NPIs with targeted antiviral policy are less effective in reducing the overall attack rate,

compared to frontline countries above. Pervasive antiviral prophylaxis for 90% of cases, their household, school, and work contacts – involving prompt contact tracing and treatment within 1 day of symptoms of the case patient – substantially reduces the attack rate even at high levels of infectiousness (R_0=2). However, such a policy is unfeasible even in well-prepared countries in conditions of sustained community transmission. A child-first pre-vaccination strategy prioritizing 0-16 year olds with a pre-pandemic vaccine (i.e., offering lower efficacy) in combination with antiviral prophylaxis for 90% cases and close contacts is effective in mitigating the U.S. pandemic. This effectiveness rises if NPIs are added to the strategy. In this study, pre-vaccination policies become more feasible and effective since vaccines can be better matched to the circulating strain and subsequently rapidly produced, with the increase in the delays in importation of the pandemic from abroad (e.g., via 99% effective border controls in the U.S.).

Another modeling study of the U.S. (Germann *et al.* 2006)[38] evaluates pre-vaccination policies – termed 'dynamic' vaccination in the study – that ramp up to cover most of the susceptible population (i.e., those still uninfected) such that there is substantial protection conferred at the point a pandemic begins to spread within the country. They find a child-first pre-vaccination policy is effective in reducing the overall attack rate even at infectiousness as high as R_0=1.9. Dose requirements for the pre-pandemic vaccine can dramatically influence the effectiveness of the policy since they change the size of the covered population for fixed resource availability of antigen. Therefore, a pre-pandemic vaccine requiring two doses is substantially less effective than one requiring a single dose. Given a dosage requirement, the effectiveness of the pre-vaccination policy rises dramatically when combined in a strategy adding TAP and NPIs. For example, a strategy combining random – as compared to child-first – pre-vaccination and NPIs nearly eliminates projected disease incidence in their model even at R_0=1.9. The authors observe that delivering a strategy combining NPIs (and/or TAP) and vaccination could be onerous for local governments, even if there are large synergies in terms of delaying the spread of the disease within the country such that more individuals are eligible for the vaccine.

The recommendations of the two modeling studies above that consider mixed strategies are contradicted by studies that consider NPIs in isolation. Of these, Wu *et al.* (2006)[27] also assess the feasibility of these NPIs in terms of the demands on local governments and the social and economic burden imposed (via the proxy of the size of peak quarantined population). The study finds that NPIs – even with conservative assumptions

on the proportion of cases reported, and compliance with subsequent case isolation or local quarantines – can halve the overall attack rate at R_0=1.8. This result is obtained with the assumption of 30% asymptomatic transmission of pandemic influenza, usually a burden on receiving benefits from contact tracing and early isolation of cases. The use of antivirals can be considered to achieve similar mitigation results – if large stockpiles are available – in lieu of isolation of cases and contacts.

Global Policy Implications

Current policy orientation in secondary risk countries remains focused on mitigation, with the objective to lower the overall attack rate while achieving delays in the peak of the epidemic such that a well-matched vaccine strategy could become feasible and local health systems are not overwhelmed. In countries where resources allow it, the strategy emphasizes antiviral prophylaxis for cases and contacts during community transmission alongside nonpharmaceutical measures as appropriate. The efficacy of antivirals for a future pandemic strain may vary from the known parameters today, and a well-matched vaccine will only be ready with a lag since the beginning of a global pandemic. Additionally, pharmaceutical policies are resource-intensive, which implies only certain countries will be able to use them.

The current distribution of global resources indicates that risk-reduction strategies that emphasize non-pharmaceutical measures alongside techniques to reduce international spread would be valuable from the perspective of secondary risk countries (the majority of countries). However, modeling of NPIs has shown varying effectiveness according to modeled context and historical evidence from the 1918 pandemic. Globally, NPIs have a role in *mitigation* at moderate R_0 and θ, even without antiviral prophylaxis or vaccination in tandem. The timing requirements for NPIs are not more onerous than for drug-based strategies, but they may have significant societal costs. The NPI effectiveness seen in the studies for S.E. Asia is higher than in studies for the U.S., possibly due to model-related factors. The early use of NPI intervention, limited seeding, and the lack of reintroductions in the S.E. Asian models may explain the observed difference. In contrast, for U.S.-specific models, given likely delays in starting interventions, reintroduction risk alongside multiple infection seeds, a persistent decline in susceptibility and infectiousness through pre-vaccination with a pre-pandemic vaccine and/or antiviral prophylaxis is preferred.

However, pharmaceutical policies are resource and capacity intensive, and may be overkill in conditions of moderate R_0 if a strategy of cooperative IATRs mixed with border control/airport quarantine in secondary risk countries is probable and effective in reducing infection seeds. Also, if containment in frontline countries following the WHO Interim Protocol[1] is progressively successful, and more countries cooperate as seen in a recent study[80] to share antiviral stockpiles, the reintroduction risk globally would fall. In this situation, for countries at secondary risk the policy objective will resemble early containment. This implies that the choice of stringent, layered NPIs at airports and reactive NPIs in communities could minimize the attack rate in secondary risk countries, and be less costly from a social perspective. Backward induction from this insight implies that if international cooperation can be ensured, the pharmaceutical-focused strategy in certain countries should be recalculated.

In situations of higher infectivity (R_0), the primary benefit of NPIs will be to delay peak transmission such that a well-matched vaccination policy becomes viable. However, any such vaccination policy that succeeds internationally will require a plan for the distribution of antigen stocks to resource-poor settings to mitigate the risk of future outbreaks as well as to meet current ethical standards. Early estimation of the infectiousness of the pandemic strain and timely and complete sharing of viral data will be essential in setting international priorities and control strategies.

Note: For specifics on the particular dose and distribution choices for vaccination strategies, see the main text. Also refer main text for details on ethical considerations of policies as well as international cooperation on viral data; as well as cost-effectiveness studies of particular control strategies.

Recommendations to policymakers

The best course of public health action for resource-poor countries to take when faced with a suspected pandemic influenza outbreak has been outlined by the WHO in its related *Protocol[1].* This study has reviewed the basis for the recommendations included in this protocol, beginning from the assumptions on the nature of a future pandemic-capable influenza strain related to its control, and the results of modeling studies that looked at several types of mixed strategies or single interventions. These results have shown a sensitivity to the threshold values of R_0, which is a modeling construct, but can be of immense value in getting a snapshot of the policy problem before the problem becomes more significant. As a result, it is recommended that the frontline risk country that is hit with an outbreak also attempt to determine the basic reproductive number R_0 rapidly from

interviews with the index cases and their contacts in order to establish a chain of infection. Averaged over various index cases and contexts, this gives a rough, ready-to-use estimate of the R_0 of the strain in circulation and sets the context for the resource investment to follow in control. Other countries which may be at risk based on their connectivity to the country with the outbreaks can then take decisions on travel bans or other restrictions based on objective criteria of risk – this is important given the economic costs of such restrictions.

Countries which are at secondary risk and have not had a related WHO recommended policy statement for 'mitigation' need to analyze their risk based on various scenarios of pandemic influenza emergence and spread to their particular shores. Their connectivity to various pandemic influenza hotspots via air networks may be one such aspect of risk, after which must be added the likelihood of spread from airports and dense urban areas around them to other cities and rural populations. These studies of risk and the likely impact of pandemic influenza in their own countries may be conducted with several levels of modeling that do not approach the complexity of stochastic-spatial modeling, and yet yield results which are better than the application of flat countrywide attack rates.

Finally, all countries at risk for pandemic influenza as well as future emergent, highly communicable diseases are well advised to analyze their private benefits from international cooperation in the sphere of infectious disease control and risk abatement. This is especially so when the cooperation requires a material contribution. This study has uncovered the potential skew in preparedness spending for pandemic influenza that does not match the distribution of risk. From 2004-07, the U.S. spent US$6.3 billion dollars on domestic pandemic influenza preparedness[81]. In contrast, the rest of the world spent approximately $1.64 billion[82]. The U.S. government should consider commissioning studies to analyze the reduction in the domestic risk of disease spread (which begins with the reduced likelihood of the disease reaching the U.S shores if more money is spent abroad). This 'preparedness at a distance' approach can be applied when the next emergent infectious disease risk is sighted. Studies in this context can closely follow the model set by Colizza *et al.*[80] The greater material contribution of the developed nations to preparedness and health systems development in the resource-poor world targeting new infectious disease – as has been seen for other diseases such as HIV/AIDS – may also ensure that when frontline risk countries must share vital disease-related genetic information, there will be little reluctance.

Cited References and Notes

1 World Health Organization 2007a "Interim protocol: Rapid operations to contain the initial emergence of pandemic influenza" Geneva: WHO

2 CDC 2004 "Influenza viruses" Online: www.cdc.gov/flu/avian/gen-info/flu-viruses.htm (6/27/07)

3 Longini *et al.* 2005 "Containing pandemic influenza at source" *Science* **309**(12):1083-1087

4 Risk Management Solutions 2007 "Analytical methods for modeling pandemic flu" Presentation at The World Bank, Washington D.C., April 25, 2007

5 Ferguson *et al.* 2005 "Strategies for containing an emerging influenza pandemic in Southeast Asia" *Nature* **437**(8):209-214

6 World Health Organization 2007b "Communicable disease surveillance and response: avian influenza data" Online at: www.wpro.who.int/sites/csr/data/data_Graphs.htm (6/27/07)

7 Centers for Disease Control 2007 "Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States" Atlanta: CDC

8 McKibbin *et al.* 2006 "Global macroeconomic consequences of pandemic influenza" Sydney: Lowy Institute for International Policy

9 Meltzer *et al.* 1999 "The economic impact of pandemic influenza in the United States: priorities for intervention" *Em Inf Dis* **5**(5):659-671

10 Taubenberger *et al.* 2005 "Characterization of the 1918 influenza virus polymerase genes" *Nature* **437**:889-893; Gibbs *et al.* 2006 "Was the 1918 pandemic caused by a bird flu?" *Nature* **440**:E8

11 Murray *et al.* 2006 "Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-1920 pandemic: a quantitative analysis" *Lancet* **368**: 2211-2218

12 Afkhami, A. 2003 "Compromised constitutions: the Iranian experience with the 1918 influenza pandemic" *Bull Hist Med* **77**:367:392

13 Centers for Disease Control 2006 "Interim guidance on planning for the use of surgical masks and respirators in health care setting during an influenza pandemic" Atlanta: CDC

14 Fraser *et al.* 2004 "Factors that make an infectious disease outbreak controllable" *PNAS* **101**(16):6146-6151

15 Bell *et al.* (WHO Writing Group) 2006a "Nonpharmaceutical interventions for pandemic flu, international measures" *Em Inf Dis* **12**(1):81-87

16 Sertsou *et al*. 2006 "Key transmission parameters of an institutional outbreak during the 1918 influenza pandemic estimated by mathematical modeling" *Th Bio & Med Modelling* **3**:38

17 Eichner *et al.* 2004 "Transmission potential of smallpox: estimates based on detailed data from an outbreak" *Am J Epidemiol* **158**:110-117

18 Pitman *et al.* 2005 "Entry screening for severe acute respiratory syndrome (SARS) or influenza: policy evaluation" *BMJ* **331**:1242-1243

19 Dept. of Health & Human Services 2006 "Pandemic planning update III" Washington DC: DHHS

20 World Health Organization 2007c "Questions and answers on pandemic influenza vaccine" Online at: www.who.int/immunization/newsroom/PI_QAs/en/print.html (6/27/07)

21 Food & Drug Administration 2007 "FDA approves first U.S. vaccine for humans against the avian influenza virus H5N1" Online at: www.fda.gov/bbs/topics/NEWS/2007/NEW01611.html (6/27/07)

22 Bansal *et al.* 2006 "Comparative analysis of influenza vaccination programs" *PLoS Med* **3**(10):e387

23 Treanor *et al.* 2006 "Safety and immunogenicity of an inactivated subvirion influenza A (H5N1) vaccine" *NEJM* **354**(13):1343-1351

24 Riley *et al.* 2007 "Optimizing the dose of pre-pandemic influenza vaccines to reduce the infection attack rate" *PLoS Med* **4**(6):e218

25 Ong *et al.* 2007 "John F. Enders lecture 2006: antivirals for influenza" *J of Inf Dis* **196**(2)

26 World Health Organization 2006a "Rapid advice guidelines on pharmacological management of humans infected with avian influenza A (H5N1) virus" Geneva: WHO

27 Wu *et al.* 2006 "Reducing the impact of the next influenza pandemic using household-based public health interventions" *PLoS Med* **3**(9):e361

28 Tellier, R. 2006 "Review of aerosol transmission of influenza A virus" *Em Inf Dis* **12**(11):1657-1662

[29] Li *et al.* 2006 "*In vivo* protective performance of N95 respirator and surgical facemask" *Am J Industrial Med* **49**:1056-1065

[30] Toner, E. 2007 "CBN report: surgical masks may provide significant aerosol protection". Online at www.upmc-cbn.org/report_archive/2007/02_February_2007/cbnreport_02152007.html (6/27/07)

[31] Rvachev *et al.* 1985 "A mathematical model for the global spread of influenza" *Math Bios* **75**(3)

[32] Stated in another way, herd immunity/effects are the double whammy results of containment policies: antiviral prophylaxis, vaccination, social distancing, etc. People not directly targeted by the intervention are also protected because chains in transmission are broken. For example, vaccinating a high-risk individual also reduces the infectiousness of their close contacts.

[33] The other is mortality: the more lethal the virus the less an epidemic can spread. Severe complications from the disease – by rendering a person bed-ridden – can do the same.

[34] Bootsma *et al.* 2007 "The effect of public health measures on the 1918 influenza pandemic in U.S. cities" *PNAS* **104**(18):7588-7593

[35] Longini *et al.* derived these probabilities based on prior studies of mixing groups from Thailand. If the distribution algorithm has assigned n individuals to that household group, then the average number of effective contacts per infected introduction = $c(n-1)$, where c: 'mixing group specific probability of effective contact per day' age stratified in a household, e.g., child-child, adult-child.

[36] The probability of infection for a susceptible per day is $p=cx$ (x: per contact transmissibility)*.* For no intervention, a school child is exposed to a number of child and adult infectives in her household, I_{hc} and I_{ha}; her neighborhood cluster, I_{nc} and I_{na}; her school I_{es}; and the two social groups she mixes in, I_{s1} and I_{s2}; with corresponding transmission probabilities for each contact of p_{hcc} (child to child), p_{hac} (adult to child), p_{ncc}, p_{nac}, p_{es}, p_{s1} & p_{s2}, respectively. The probability P the child becomes infected on the day = $1 - (1 - p_{hcc})^{Ihc}(1 - p_{hac})^{Iha}(1 - p_{ncc})^{Inc}(1 - p_{nac})^{Ina}(1 - p_{es})^{Ies}(1 - p_{s1})^{Is1}(1 - p_{s2})^{Is2}$

[37] Ferguson *et al.* 2006 "Strategies for mitigating an influenza pandemic" *Nature* **442**(27):448-452

[38] Germann *et al.* 2006 "Mitigation strategies for pandemic influenza in the United States" *PNAS* **103**(15):5935-5940

[39] Doyle *et al.* 2006 "Influenza pandemic preparedness in France: modelling the impact of interventions" *J. Epidemiol. Community Health* **60**:399-404

[40] They use GAR not in the sense of infections (exposures) but the number of clinical cases caused by pandemic flu per unit population during a wave.

[41] Example – if influenza-related death rates among 20-64 yr olds range between 0.06-0.15 per 1000 persons given a GAR of 15%, does it make economic 'sense' (return a positive net present value) to vaccinate everybody in this age group? They use two scenarios of base vaccine efficacy.

[42] van Genugten *et al.* 2003 "Pandemic influenza and healthcare demand in the Netherlands: scenario analysis" *Em Inf Dis* **9**(5)

[43] Institute of Medicine 2006 *Modeling Community Containment for Pandemic Influenza - A Letter Report* Washington DC: The National Academies Press

[44] If $\rho_{.0}$ and $\rho_{.1}$ denote the transmission probability to unvaccinated and vaccinated susceptibles, respectively, where the dot in the subscript can denote any vaccine status or an average across the population, then in $VE_s=(1- \rho_{.0}/\rho_{.1})$, VE_s represents the proportionate reduction in the transmission probability given a specified exposure to infection. Analogously, if $\rho_{0.}$ and $\rho_{1.}$ denote the transmission probability from unvaccinated and vaccinated exposed to susceptibles, respectively, where the dot in the subscript can denote unvaccinated susceptibles or an average across the population, then in $VE_I=(1- \rho_{0.}/\rho_{1.})$, VE_I represents the proportionate reduction in the infectiousness of the vaccinated. See: Halloran *et al.* 1999 "Design and interpretation of vaccine field studies" *Ep Rev Vaccines* **21**(73)

[45] Longini *et al.* 2004 "Containing pandemic influenza with antiviral agents" *Am J of Epidemiol* **159**(7)

[46] See discussion of such effects in the context of vaccines in Halloran *et al.* 1999

[47] The assumed rapid spread within a country depends on internal connectivity. In a large country such as the US with sparsely populated areas, but more domestic long- and short-distance travel, differences in attack rates across locations may be equalized as rapidly as 1-2 weeks after the first outbreak, as per Wu *et al.* (references therein). This may not follow in other country contexts.

[48] This requirement may inflate if the possibility of false positives in case identification – given that symptomatic diagnosis of influenza viral infection is often inaccurate – is considered.

[49] The international quality and currency of outbreak information has improved, based on changes since SARS. The WHO official and rumor surveillance networks, along with the resources of the CDC could ensure that the delays in receiving information may be minor.

[50] Oldham, J. 2005 "LAX plans for bird flu quarantines", The LA Times, October 18

[51] CDC 2003 "Revised U.S. surveillance case definition for Severe Acute Respiratory Syndrome (SARS) and update on SARS Cases - United States and Worldwide, December 2003", *MMWR* **52**(49)

[52] Dept. of Health and Human Services 2005 *HHS Pandemic Influenza Plan* Washington DC: DHHS

[53] The simplest model in Riley *et al.* (2007) is: $a=(1 - p_v*c)(1 - e^{-R0*a})$ where a is the attack rate, c is the proportion of the population vaccinated, and p_v is the proportion of c who receive complete protection. The first part of the equation describes the initial proportion susceptible and the latter is the overall probability of infection. If p_v is 39% with two 10μg doses, then if 20 out of 300 million people in the US are vaccinated, the initially susceptible proportion $= 1 – (20/300)*0.39$, and the attack rate drops from the no-vaccination baseline of 73% (at $R_0=1.8$) to 69.5%.

[54] Fraser, C. 2007 "Influenza pandemic vaccines: spread them thin?" *PLoS Med* **4**(6):e228

[55] Bell *et al.* (WHO Writing Group) 2006b "Nonpharmaceutical interventions for pandemic flu, national and community measures" *Em Inf Dis* **12**(1):88-94

[56] Hatchett *et al.* 2007 "Public health interventions and epidemic intensity during the 1918 influenza pandemic" *PNAS* **104**(18):7582-7587

[57] In Hatchett *et al.*, 'timing' of policies could explain only 50% of the variation in total excess mortality across their sample of cities, an indicator closely tied to the CFR. Unknown factors that could account for variation in CFR may include: differences in general public health (which may also tie with observed differences in mortality in Afkhami 2003), varying levels of bacteria in the environment, etc. The authors rule out antigenic variation in the viruses across waves as a factor.

[58] Carrat *et al.* 2006 "A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics" *BMC Med* **4**(26)

[59] See Wu *et al.* for a more involved graphical depiction of a stochastic SEIR model.

[60] Wu *et al.* reviewed the available studies on θ and found the range to be 0-50% (based on Rvachev *et al*). Their assumed value is then slightly above the median.

[61] Brownstein *et al.* 2006 "Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States" *PLoS Med* **3**:e401; Viboud *et al.* 2006 "Air travel and the spread of influenza: important caveats" *PLoS Med* **3**(11):e503

[62] Cooper *et al.* 2006 "Delaying the international spread of pandemic influenza" *PLoS Med* **3**(6): e212

[63] Epstein *et al.* 2007 "Controlling pandemic flu: the value of international air travel restrictions" *PLoS One* **2**(5):e401

[64] Epstein *et al.* conclude that pandemics that originate in January in Hong Kong could, due to effective international air travel restrictions, lead to a peak in the temperate Northern hemisphere in September. Therefore, the delay of epidemic introduction would push major transmission into the influenza high season, leading to more cases. This assumes that pandemic influenza A would demonstrate similar seasonality in their infectivity as interpandemic strains. This effect would be reversed for a July start to the pandemic in the tropics, which would cause a peak in spring.

[65] Jani, M. 2005 "Modeling the large scale disruptions of an airline network" *J Trans Engg* **131**(4): 249-260. For more on the hub-and-spoke nature of the international network in the context of communicable disease spread, see Colizza *et al.* 2006 "The role of the airline transportation network in the prediction and predictability of global epidemics" *PNAS* **103**(7):2015-2020

[66] Pandemics of severity category 1-5, based on their CFR. See CDC 2007, Table A, pp. 12

[67] Susarla, A. 1996 "Plague in Surat, India" Unpublished. Clark University Gr. School of Geography.

[68] Hai *et al.* 2004 "The short-term impact of SARS on the Chinese economy" *Asian Econ Papers* **3**(1):57-61

[69] Lee *et al.* 2005 "The impact of the SARS epidemic in Taiwan: implications for human resources, labour markets, and unemployment in the service sector", Working Paper Series, Judge Business School, University of Cambridge

[70] Siu *et al.* 2004 "Economic impact of SARS: the case of Hong Kong" *Asian Econ Papers* **3**(1):62-83

[71] Brahmbhatt, M., Dutta, A. 2008 "On SARS type economic effects during infectious disease outbreaks," Policy Research Working Paper #4466, The World Bank: Washington, DC.

[72] Emanuel *et al.* 2006 "Who should get influenza vaccine when not all can?" *Science* **312**:854-8555; also see *Letters* in the same issue critiquing Emanuel *et al.* (from Frey, H., and Galvani *et al.*)

[73] World Health Organization 2006b "Global consultation on addressing ethical issues in pandemic influenza planning: summary of discussions" Geneva: WHO

[74] Gostin, L. 2006 "Public health strategies for pandemic influenza: ethics and the law" *JAMA* **295**(14):1700-1703

[75] Huang, Y. 2004 "The SARS epidemic and its aftermath in China: a political perspective" in *Learning from SARS: Preparing for the Next Disease Outbreak* Washington DC: National Academies Press

[76] Fidler, D. 2004 "SARS: political pathology of the first post-Westphalian pathogen" in *Learning from SARS: Preparing for the Next Disease Outbreak* Washington DC: National Academies Press

[77] World Health Organization 2006c "Avian and pandemic influenza: developments, response and follow-up, and application of the International Health Regulations" Geneva: WHO

[78] World Health Organization 2007d "Best practice for sharing influenza viruses and sequence data" Report by the WHO Secretariat to the Executive Board, Geneva: WHO

[79] For a true or 'pure' public good, it would be necessary that it possess a 'non-excludable' nature, i.e., it would be too expensive to exclude anyone from using it; and is 'non-rival' in consumption, i.e., use by one person would not diminish its availability for another.

[80] Colizza *et al*. 2007 "Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions" *PLoS Med* 4:e13.

[81] Lister, S.A. 2007 "Pandemic influenza: appropriations for public health preparedness and response" *CRS Reports for Congress*, Congressional Research Service. Washington D.C.: Library of Congress

[82] UN System Influenza Coordinator & The World Bank. 2007. "Responses to avian influenza and state of pandemic readiness: third global progress report" December 2007. Washington D.C.: The World Bank.

Paper II: Potential Infections during a Pandemic of Influenza in the States of India: Implications for Public Health Policy and Research

Summary

Background: In a future where pandemic influenza outbreaks occur globally, eventual transmission of the disease to India is highly likely. Recent modeling studies indicate that delay of the peak of the country epidemic can be achieved by effective controls on air travel. This study extends the analysis of such benefits of costly, proactive interventions for suspension of international travel in the Indian context by connecting them to eventual reductions in attack rates. These additional benefits are realized because internal spread of pandemic influenza under the scenario of 'air transmission' is higher than a scenario where the disease enters the country across land and sea borders later in the pandemic timeline.

Methods: A methodological review of classical SEIR and spatial models is provided as a basis for the overall approach. Age-structured average attack rates of pandemic influenza calibrated to past pandemics and inter-pandemic influenza are available from the literature. Using these attack rates, internal spread of pandemic influenza in India is modeled using scenarios of disease transmission to India and using weights to scale the attack rates faced by specific provinces under the scenarios. Provinces face a higher attack rate based on the location of international entry points (air, land, sea) on their territory, the internal transport networks, and their demographics. The results from the scenario-based approach for New Delhi are also compared to the epidemic size from a structured metapopulation experiment in EpiFlex software. The overall attack rates in India from the scenario-based analysis are used to scale estimates of averted health and economic costs from prior studies using a fixed attack rate, to yield a range of benefits from the proactive closure of air travel.

Findings: The range for potential pandemic influenza epidemic size in India at a reference value of the basic reproductive number, R_0=1.4, is 7%-21% of the population, depending on the transmission scenario and assumptions. This is below the widely used 30% gross attack rate figure in the literature. The ranges of additional benefits from preventing transmission of disease by air are 61,000-100,000 excess fatalities and US\$400-710 million, depending on the assumptions. The epidemic size and the R_0 from the metapopulation model are higher than the scenario-based model, indicating that even improved average-based influenza models as the latter may underestimate individual-level disease propagation factors, such as superspreading. Results of the EpiFlex experiment show that superspreading may be an important element driving rapid and large epidemics in megacities such as New Delhi.

Interpretation: This study is intended as an aid for public health decision-making under conditions of substantial uncertainty. Based on constantly updated risk of a pandemic of influenza transmission to India, the Government of India should enact proactive and complete suspensions of air travel to the country. Though further research is required, the social benefits may outweigh the private costs. Further research is necessary to model individual-level factors responsible for influenza propagation in India's cities.

Introduction

Human cases of avian influenza A of subtype H5N1 have raised the possibility that the virus could mutate and cause a human pandemic. A pandemic-capable virus would be more infectious in humans compared to the avian subtype H5N1. The latter has a case fatality rate of 61.5% based on 346 cases and 213 deaths across fourteen countries since 2003 (1). The pandemic-capable virus that emerges would have lower virulence, but how much lower is an open question. The infectiousness and virulence are set in the process of mutations in the virus as it adapts to its human host[7]. The number of mutations required for influenza virus strains adapted to birds to become pandemic-capable in humans may be only two (2). Alternatively, a pandemic strain may emerge via a reassortment of viral genes in a host co-infected with seasonal human influenza and avian influenza A of subtype H5N1 (as in 1957 or 1968, which were less virulent strains). The probability of such reassortment leading to a pandemic-capable adaptive mutation generally rises with increased geographical spread and frequency of human cases of avian influenza A.

Given the difficulty in predicting the timing or results of the adaptive mutation event, the impact of a worldwide pandemic is usually not estimated as an expected value (however, see the *Discussion* section below). Instead, 'if-then' analyses are common, where the emergence of a pandemic strain in humans is taken as a given. In the world where successive outbreaks[8] of a pandemic-capable form of influenza A disease occur internationally, it would be almost impossible to prevent the transmission of the disease to India, but transmission may be delayed, with large public health benefits. These benefits are the focus of this study. In the state of the world where several geographically distributed outbreaks of pandemic-capable influenza have already occurred, the transmission of the

---

[7] Influenza infectiousness between mammals depends on whether the virus strain will attach easily to cells in the upper respiratory tract (nose and throat). But, the virus will be more virulent if it binds and replicates in the lower respiratory tract (lungs). Experiments in a ferret model have also shown that infectious viruses provoke sneezing in the nose, which boosts contagion (2). While the avian H5N1 subtype shows a preference for binding and replicating in the lungs (poor infectivity, high virulence), most sub-strains of the 1918 avian-derived pandemic virus preferred to bind in the nose/throat but also replicated in the lungs – mixing high infectiousness with virulence (2).

[8] The distinction between 'outbreak' and 'epidemic' is nuanced. Some sources consider them synonymous (3), but in this study, and in the context of a highly infectious disease, an *outbreak* is a sequence of related infections that occur in a short period of time in a localized and defined area smaller than a city (e.g., neighborhood, airport). An *epidemic* is a sequence of outbreaks across a city or set of cities; or the higher level of a single outbreak that crosses a predefined threshold set in terms of geographical coverage (citywide, statewide, or nationwide) or in terms of number of cases.

disease to India can be taken as a 'high probability, high impact' public health event. For evaluating interventions to create such delay, it is of interest to policymakers to estimate the potential size of the impact of a pandemic under various scenarios of transmission. Such a calculation is necessary to rationalize potentially costly prevention measures such as closure of borders.

Outbreaks in India: The assumption of a high probability of pandemic transmission to India is based on the relative difficulty in preventing the arrival of infected individuals across air/land/sea borders in the context of rapid worldwide dispersal of infectives from the original outbreak points. This is based on the prior pandemics, as well as recent work analyzing the worldwide transmission through airline networks of SARS in 2003 (4) and seasonal influenza (5, 6). Models that analyze a WHO-sponsored policy of reactive air travel suspension imposed in the wake of the initial outbreaks, indicate that these measures may have limited effect on the overall number of infections (7, 8) though they may delay the peak of the epidemic in participating countries (9), given a pandemic influenza strain of moderate infectivity. The analysis below will suggest that this underestimates the benefits of such air travel suspension, given some assumptions on the intervention efficacy.

There may be several scenarios under which outbreaks occur in India. In order to construct the scenarios, it is useful to review how transmission to the country may occur. In each scenario, transmission to India occurs with a high probability over the entire course of a wave of the global pandemic, but what differs is the interval between the first foreign outbreaks to the first Indian outbreak. In the first scenario of 'airborne transmission', the delay in occurrence of outbreaks in India after initial worldwide outbreaks – determining the preparedness of the Indian public health system to the pandemic challenge – will depend on the location of such initial outbreaks. In the unlikely event that rapid suspension of air travel ordered by the Government of India prevents all airborne transmission, outbreaks in India may occur later in the course of the worldwide pandemic, after infected individuals arrive in the country across land or sea entry points, especially once neighboring nations have a high prevalence of the disease. This scenario is termed 'land/sea-based transmission'.

The significance of the Indian outbreaks for the ultimate size of local epidemics, given a scenario of air-borne transmission, crucially depends on the choice of location for the first international outbreaks. For example, if the first outbreaks occur in Singapore,

there may be little advance warning for public health preparedness given the frequency of flights to multiple Indian destinations from that city and hence the volume of potential infectives entering India. In other words, the cost is high for any confirmatory delay from the moment the nature of the outbreak in Singapore is revealed, till the Government of India decides to suspend air travel to and from that city state (or a broad definition of risky destinations). A policy to suspend air travel, given its economic and social costs – such as the stranding of Indian passengers in a foreign location – will be implemented with due review. This implies that if the first international outbreaks occur in Singapore, local public health authorities in India will have little time to prepare before outbreaks occur and this will lead to larger epidemics, if all else is constant.

However, the potential range of initial outbreak locations and hence the range of warning times for public health preparedness in India is large. It is erroneous to assume that the first pandemic influenza outbreaks must occur in areas with currently high levels of human cases, e.g., Indonesia. The adaptive mutation event is also possible in Europe and Africa, given the changes in currently circulating avian influenza viruses in those continents (10). If the first outbreaks occur in Africa (e.g., in countries with prior human cases, such as Egypt), the low frequency of flights from these countries to India implies a lower total cost of the confirmatory delay till air travel from the risky destinations is suspended. The converse is true if the first outbreaks occur in Europe or Asia, especially near major air travel hubs of significance to India, such as London, Manchester, Singapore, or Hong Kong. Later in this study, the potential range of initial locations of outbreaks within the air-transmission scenario will be captured by varying the strength of local Indian preparedness.

The airborne transmission and the land/sea based transmission scenarios differ on the important note of the local geography of the Indian outbreaks. In the airborne transmission scenario, initial outbreaks in India occur in large metropolitan areas that feature an international airport. These areas are also densely populated and are well connected in a domestic transport network. In the land/sea based transmission scenario, outbreaks occur in port cities and border areas, of which the latter imply lower population densities and transport connections. Also, a significant interval since initial warning may have been achieved, implying local public health systems may be better prepared than in the first scenario. Therefore, holding all other things constant, we would expect worse local (provincial) and national epidemics with the first scenario than the second.

In the third and Worst Case scenario, the outbreaks occur in a highly networked location in Southeast Asia and then spread in the region neighboring India, and the government is unable to prevent the entry of significant numbers of infectives via air, sea, and land routes, such that many more local outbreaks are sparked. The outbreaks caused via the air route occur without any delay from the occurrence of the global outbreaks, while the subsequent outbreaks due entry of infectives through the sea and land routes occur after a short interval. Therefore, the third scenario can be taken as a sum of the effects of the first two scenarios. In all, only three scenarios are selected for analytical simplicity, spanning possibilities in the international emergence and transmission of pandemic influenza, as well as the uncertainty in policy effectiveness for preparedness.

Outbreaks to Local Epidemics: As they are constructed, the first two scenarios associate different risks of large epidemics for the individual states of India (provinces). This implies that each scenario involves a different projection of total infections in India during one wave of the pandemic. This difference in risk has three levels. If all other things remain constant, the differences in the size of the resulting epidemics across the states of India in a scenario will be driven by their exposure to the risk of hosting the first domestic outbreaks. Those states with the first domestic outbreaks (e.g., from international air travelers) will have less time to prepare health systems, and will suffer greater infections, ceteris paribus. This *primary risk* differential across the states derives from the presence or absence of international airports, international sea ports, or land border crossing points on the territory of the state. Given the primary risk, the *secondary risk* differential across states derives from the dynamics of how outbreaks spread and become epidemics, involving local demographics, density of population, and transport networks. Finally, a *tertiary level* of risk differential relates to variance in local public health system capacity across states, especially for enacting the more feasible, nonpharmaceutical interventions (see Table 2 below), which impacts the state levels of infections during a countrywide pandemic influenza wave.

The scenarios above describe a framework to estimate the national total of infections during a wave of pandemic influenza in India in which three state levels of risk factors are considered: transport links, demographics, and local health system capacity. The third section below (*Methods Used*) discusses how these weights are constructed and applied, and how the process relates to insights from the literature and to the overall policy question. The methodology of the weighted scenarios is intended to be a policy-relevant,

realistic yet analytically tractable process to describe the local epidemics of pandemic influenza which might emerge from asynchronous outbreaks in different parts of India.

Modeling Epidemics: There are several epidemiological techniques available in the literature that model influenza epidemics, suited for different policy questions. Some ways of modeling an epidemic are suited to testing public health interventions that control and prevent outbreaks from spreading in specific geographic and demographic conditions. The progression from outbreak to epidemic can be simulated in a way such that factors that affect person-to-person transmission are included and policy interventions that modify those factors can be tested. Other ways of modeling are suited to estimate the epidemic curve (number of infections per day, over time) in a defined location where factors affecting person-to-person transmission are constant over time and across persons. Such methods yield the likely timing and size of the peak rate of cases per day under different initial conditions and assumptions, and can aid planning for public health preparedness, such as estimating the required surge capacity. The need for data and computational resources varies across these epidemiological modeling methods.

The stated policy intent of this study is to estimate the potential number of infections at the national and state level under several scenarios of transmission to India, in order to yield insights relevant to resource allocation. Some studies have provided estimates of the total size of an epidemic across India – one of many countries included in their scope – under several scenarios of infectiousness and virulence of a pandemic influenza virus (11, 12). These studies do not conduct epidemiological modeling given their indicative purpose and the inherent tradeoff between model realism and model parsimony. Even with the imposition of this tradeoff, several improvements are possible. Improvements as attempted in this study are described in a methods section further below.

Other studies with actual epidemiological models have tested control interventions for pandemic influenza outbreaks in Southeast Asia (13, 14) and for mitigating a widespread epidemic in the U.S. (15, 16). The insights from such studies may be broadly applicable, but given the different demographics, individual and population-level conditions, transportation networks, seasonal and climatic differences, etc., relevant to disease propagation in the Indian setting, epidemiological models should be reapplied to the country's specific context. The resources and data required for such an India-specific endeavor would be considerable. This study does not intend to test specific public health

interventions suited to preventing epidemic take-off in India. However, certain insights relevant to outbreak control policy will be available from the analysis.

Plan for this Study: The *Survey of Methods* section below briefly discusses epidemiological modeling techniques for influenza commonly used in the policy literature, such as the various forms of compartmental (SIR or SEIR), spatial-stochastic, and scenario-based models. The *Methods Used* section synthesizes the survey of methods to argue for and describe a specific methodology chosen in this study to execute the scenarios. This methodology is based on insights from the literature, but tailored to the policy question and the available resources. Data sources, formulae used to construct the risk weights, and specific tables of weights by state are presented in the supplementary materials (appendix). The *Results* section reports the state and national numbers of infection under the different scenarios, and with the inclusion of the geographic/demographic and health policy weights. Sensitivity analyses to the value of the weights used are presented in the supplementary materials. Additionally, the results of an in-silico experiment of a pandemic influenza epidemic in a metapopulation scale model of New Delhi city – executed in EpiFlex software – are also presented. This experiment is a demonstrator of spatial modeling techniques that could be applied to the Indian national level with greater resources, as well as to produce an EpiFlex-derived attack rate number that could be compared with the value from the execution of the weighted scenarios for New Delhi state.

The *Discussion* section has two parts which synthesize the results and put them in policy and research context. The first part examines the results in terms of the averted epidemic size in India under various scenarios in order to understand the benefits of costly outbreak prevention interventions. The second part compares the scenario-based and EpiFlex results on policy implications and on methodological standpoints. In this context, plans are proposed for using structured metapopulation modeling using platforms such as EpiFlex to better understand infectious disease epidemic dynamics in India.

Survey of Methods to Model Influenza Epidemics

Given an outbreak, how can one predict the overall number of infections over the course of a resultant epidemic in the surrounding area, during which pandemic influenza transmits from person-to-person? The cause of an epidemic ending is usually 'herd

immunity'. In the classical sense, this is reached with the decline in the number of susceptible individuals below a sustaining threshold because most people have already been infected, and have subsequently either died or recovered with immunity. In the network concept of disease propagation, the increase in the number of immune individuals decreases the possibility of contact between infected and susceptible persons; thus decreasing the rate of spread of disease and bringing about 'herd immunity' (17). Some public health interventions which prevent movement of individuals or force them to stay indoors also work in the same way to interrupt the network of transmission.

The analysis of how outbreaks may lead to widespread epidemics of pandemic influenza is complex, especially if community-level transmission of the disease is expected (as distinct from transmission within healthcare settings that handle cases). This is because the biology of the actual pandemic virus, the varying response of a population to an outbreak, seasonal effects, etc., can all modify the overall number of infections and yet are difficult to predict beforehand. There are several analytical strategies that attempt to model the real situation of an epidemic. All models are abstractions of reality and have their faults. The value of an analytical model is based on its effectiveness in answering the policy questions it addresses. The sub-sections below discuss some of the modeling techniques related to an estimation of defensible values of total infections at the state and national level during a wave of pandemic influenza across India.

Non-epidemiological Methods: Non-epidemiological studies (10, 11) use historical data from the twentieth century influenza pandemics to estimate a flat rate of infections across the entire population during a wave, known as the Gross Attack Rate (GAR). This technique assumes the same risk of infection for every individual, and hence GAR is applied as a fraction of the population. The use of weights can help drive variation in the GAR across countries based on the impact of geography and transport networks (i.e., factors which lead to a larger proportion of the population coming under risk of infection) and the quality or capacity of local health systems (11). For the weights to be comparable across countries, they must be constructed from widely collected, internationally recognized variables.

In terms of the focal policy question of this study, the main disadvantage of the historical GAR method is the likelihood of over-estimation or under-estimation of the national figure for potential total infections during a wave of pandemic influenza. Hypothetically, if the virulence and infectivity of the virus that appears in the future is the

same as the viral agent in the 1918, 1957-58, and 1968-69 pandemics, then over-estimation is still possible because the historical GAR derives from community-level transmission in the context of a weaker public health response: lack of effective antiviral prophylaxis, poor knowledge about the virus, and poorer health of certain populations; all correlated with enhanced susceptibility. Similarly, under-estimation is possible because human populations in urban areas such as that of India are far more densely distributed and more highly connected through transport networks, allowing a more rapid and effective dispersal of infective individuals.

In addition, the fixed GAR approach neglects the important differences in the risk of infection across individuals because of their differing geographical disposition, membership of different age-groups, and different behaviors. Such factors affect the risk of infection inasmuch as they modify the natural susceptibility of an individual (likelihood of developing a clinical infection if exposed to the infectious agent) as well as the chance of meeting an infected individual. Given the actual size of the high and low risk pools of individuals in the nation (or a state), the result of using a fixed GAR may be over- or under-estimation compared to the true risk. Finally, if a future pandemic influenza-capable viral strain is more or less infectious than in the past while still being comparably virulent, the results of using a single, fixed historical GAR will be over- or under-estimation.

Compartment Models - SIR/SEIR Techniques: The simplest, non-stochastic models of influenza transmission – pandemic or seasonal – are based on the SIR (Susceptible, Infected, and Recovered) 'compartment' models of 1920s vintage. After the addition of the 'exposed' category, the resultant SEIR models are the root of deterministic modeling of epidemics. Fig. 1 on this page is an unorthodox representation, modified to show some, but not all, impact points of public health interventions. Individuals transition from being susceptible (S), to exposed (E, i.e., infection is latent), to infectious (I) and then to removed (R, recovered or dead) based on predetermined probabilities. Other preset parameters are the length of time individuals stay in state E without becoming infectious (latency period), and the time they stay infectious before being removed. The model is specified in differential equations that define the stock of people in the various states after a discrete time step, usually a day. At time zero, seed infections are introduced and the model is run.

Fig. 1: Simplified SEIR model diagram, with potential interventions
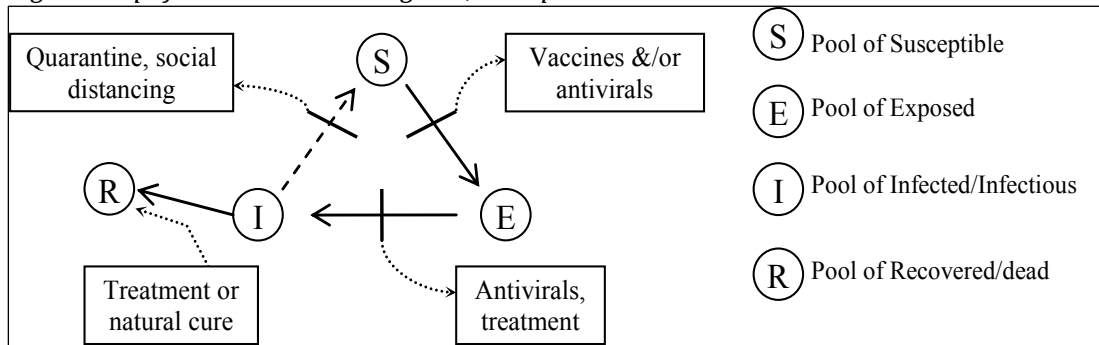
For influenza, the possibility that any infected person contacts a susceptible during a time step is set as the fraction of the population susceptible at that point, multiplied by a fixed decimal: the 'probability of infection conditional on contact'. The SEIR method has the assumption of population homogeneity with 'uniform mixing'. This implies every person is equally at risk of infection and every individual in the population is equally likely to contact someone else. The SEIR models neglect the fact that individuals do not mix uniformly; each person has contact rates depending on their social network and behavior. Such individual-level factors cause the risk of infection to vary across individuals, who can be grouped based on the intensity of such factors in different population categories, for example, urban citizens vs. those from rural, low population density areas.

The policy intent of SEIR models is to describe transmission dynamics at the population level by varying attributes of individuals that locate them within compartments (disease state, age, reported case or hospitalization status). Public health interventions – such as antiviral prophylaxis – can be introduced, and the effect observed via a change in the probability of transmission per contact across the various infective group members and the remaining susceptible (18). For capturing more of the variation in risk of infection across individuals, SEIR models have to add compartments without losing mathematical tractability. Several improvements in this vein have been made, which can be discussed by relating them to the effective reproductive number of the epidemic, or R.

In this context, the basic reproductive number, R_0, captures the *a priori* infectiousness of a communicable disease in a particular setting. It is defined as the number of secondary infections caused by a single typical infected case in a completely susceptible population, and in the absence of interventions. At the start of an epidemic, it determines how quickly the epidemic will spread. The effective reproductive number, R, is equal to R_0*s where *s* is the proportion of the at-risk population still susceptible. Theoretically, if R can be

pushed below 1, an epidemic usually dies out due to herd immunity. Compared to the *a priori* R_0, R is always lower. Like R_0, R depends on factors including:

- The risk of transmission per effective contact, ρ
- The number of effective contacts an average person has per unit time, β
- Relative duration of the latent and infectivity periods

The effective reproductive number R is an important factor in determining how many infections will be suffered in one wave of the epidemic, i.e., the gross attack rate. The other factor determining this is virulence. Infectious diseases that rapidly kill their hosts or render them immobile also reduce their own chances of spreading widely, ceteris paribus.

Population Heterogeneity and SEIR: In the no intervention case, average R may vary across population groups based on age structure and the types of social networks. The SEIR models can be improved to account for age structure (19). The importance of age as a determinant of susceptibility to a clinical case of pandemic influenza infection may be due to both behavioral and biological factors. The W-shaped shaped curve relating excess mortality rate in percent during the 1918 pandemic (y-axis) to age groups (x-axis) suggested that deaths were highest in the segments between 15-39 years and lower for higher and lower age groups (20). Mortality reflects both the underlying attack rate and the case fatality rate (CFR). There are hypotheses for the higher CFR in younger individuals, which relates to hyper-interaction of the symptoms of avian-derived pandemic influenza viruses and the more robust immune systems as present in younger persons. For this study, the implication of the W-shaped curve for the variation in the attack rate is stressed. The portion of the variation in mortality by age during 1918-19 that is explained by age-differenced attack rates is unknown (20). Estimates of age-differenced attack rates using data from later pandemics suggest that a biological factor may be in play, related to the particular influenza A subtype that causes the pandemic. The 1957-58 pandemic with subtype H2N2 displayed a higher attack rate in U.S. children, whereas the 1968-69 influenza pandemic with subtype H3N2 had a similar attack rate across age groups (14). In general, younger people shed more viral material (21). As such, the risks of infection for children in schools would be higher.

Behavior also differs across age groups in ways that affects susceptibility. In high population density areas, younger people have contacts more frequently and at closer quarters – e.g., on urban transport, in offices, and on playgrounds. Older people who are

homebound and do not have as many contacts have a lower risk of being near an infectious person, but when near such a person, may have a higher risk of getting infected because of weaker immune systems.

In SEIR models, one way to simulate these differences in behavioral and biological factors of susceptibility is to introduce varying transmission probabilities defined in advance for different age compartments. With more computing power available in recent years, such age-structured models have become more common.

When simulating the effectiveness of public health policies, as much heterogeneity in risk of infection is desirable as is reasonable to model. Location is another important source of such variation. Urban areas can be quarantined or social distancing measures imposed, which cut the contact rates in the population. But these are more difficult to impose in rural areas where there are less defined modes of entry and exit and the population is harder to reach with public health communication. Therefore, while in rural areas the average number of effective contacts per person might be low, the public health measures might have low impact. And as occurs for older people, rural people on average may have poorer health status. They may also have poor access to antivirals. In the pandemic influenza case, since the entire country is susceptible to pandemic influenza and rapid transmission via air and road networks is probable, modeling simultaneous SEIR progress in multiple cities that exchange susceptible and infectious people is desirable. However, this requires mainframe computing resources given the time step of a day.

Physical Distribution and Structure of Populations: Adding individual-level heterogeneity based on location to SEIR models increases complexity and computational resources required. Further, SEIR models cannot involve the importance of the physical distribution and the interaction structure of the population across an epidemic's geography. In contrast, three types of modeling techniques currently in use for understanding influenza epidemics focus on incorporating population structure and distribution.

First, *spatial models* (or, 'explicit' spatial models) see influenza transmission as a function of geographical distance, i.e., a 'distance kernel'. This operates on the insight that infected individuals emerging from an outbreak travel only certain maximum distances, describing a circle of risk of infection. The number of susceptibles subject to this risk and hence the epidemic potential depends on the population density in that circle. Importantly, the effect of developing symptoms on mobility (a decline), and hence on the number of

effective contacts an average person has per unit time, β, can be explicitly modeled. This leads to a more realistic overall R compared to SEIR techniques. Such distance kernels featured in recent models of pandemic influenza outbreak and spread in rural SE Asia (13, 14). The possibility of occasional long-range travel and contact (e.g., via domestic train or air travel) can be accounted for separately. In general, the more effective and rapid is the transport network – for example in urban areas of developing nations, or in developed countries – the less is the value of using a distance kernel (22).

Second, *social network analysis* – or more generally, interaction structure models – can help to characterize how influenza spreads through a group of connected individuals comprising both infectives and susceptibles (23). For example, a network may comprise an organization (school, workplace) or many organizations, homes, and other locations. These models are generally applicable to urban areas and can help understand phenomena such as super-spreaders ('nodes'/people in the network with very high 'degree'/contacts). Like explicit spatial models, the effect of symptomatic illness on mobility and hence on β can be included to reveal more realistic GAR values. The more dense and highly connected a network (e.g., a slum in India), the higher is R and more difficult is epidemic control (24). The effect of specific interventions such as vaccination can also be studied in networks. A disadvantage of the technique is that as the number of nodes and levels of interaction structure are increased, the number of free parameters and assumptions required for calculation becomes very large (22). Background information on the social network may also be difficult to obtain in some country contexts. In the case of analyzing pandemic influenza spread in a U.S. city, such information was derived from transport data (23).

Third, *metapopulation models* represent a "theoretical compromise" (22) between the compartmental techniques and the two previous spatially-defined techniques (explicit spatial and social network). Metapopulation models usually have at least two levels of population structure: the demographic group to which individuals belong and which defines their behavior, and then certain defined locations such as cinema halls, transport hubs, etc., where individuals from different groups come into contact and mix uniformly as in compartment models. These locations can in turn be grouped at another level of structure such as districts within a city (22), or a city in a network of cities within a country (25), or a city within a global network of cities (26). Metapopulation models acknowledge the insight of social network techniques that between-group interactions of individuals depend on their routine behavior such as daily movements to and from school or work, the nature of

transportation networks, and the size and density of mixing spaces. The topmost levels of structure – e.g., cities – differ from each other in their distribution of individuals in demographic groups, the size and importance of various locations, and the movement pattern of demographic groups through the locations. This implies that very unique contexts can be simulated to study the transmission of influenza, usually at lower computation cost than large social network topologies.

The current state of the art involves the use of several spatial techniques simultaneously. In a recent study of pandemic flu spread in rural SE Asia, the location contexts are called 'mixing groups', e.g., offices, schools, and households, and for each such context a specific probability of effective contact per day is defined (14). The age-distribution of individuals, number of mixing groups and their average sizes, as well as their clustering are set so as to mimic the real population of rural Thailand. Average inter-mixing group distances used to generate the spread across the simulated space as well as to initialize the distance kernel were based on GIS data. At the time of model generation, 500,000 individuals are distributed as per an algorithm in close contact mixing groups (households, schools, workplaces), as well as casual/social contact mixing groups (temples, markets, shops). The people have effective contacts based on the assumed average rates for that mixing group context. For example, the authors assume that the probability of two children making at least one effective contact per day in a household mixing group is 0.6[9].

Epidemiological Modeling – Adding Stochasticity: In spatial models, randomness is introduced in various ways to mimic the uncertainty of epidemic spread. The analysis of stochasticity also allows the construction of probability density functions for the risk of infection (27). A stochastic-spatial epidemic model simulates the epidemic process as a series of random events in space and time with the probability of specific events defined by the model parameters. The two event probabilities of interest are the probability of effective contact (c), and probability of transmission on contact (x). The former, c, varies at the individual level, based on factors such as the location, and characteristics of that location. The latter, x, varies in reality based on age, public health context, individual behavior, etc., and is usually a constant in analysis as it is difficult to model.

---

[9] The authors derived these probabilities based on prior studies of mixing groups from Thailand. If the distribution algorithm has assigned n individuals to that household group, then the average number of effective contacts per infected introduction = $c(n-1)$, where c is the 'mixing group-specific probability of effective contact per day' as stratified by age in a household: child-child, adult-child.

For example, the authors of the study on pandemic influenza in Southeast Asia (14) introduced stochasticity by conducting Bernoulli trials and generating *N* uniform [0, 1] numbers for *N* mixing groups. If the probability for an individual to be infected on a day in that group was greater than the random number, then a single infected individual in the latent period was introduced into each of *N* mixing groups at model inception. Secondary infections resulted in each mixing group that had an infected individual. Similar to 'hierarchical epidemic' models, the within mixing group dynamic is extended to between-group analysis. The spread to mixing groups lacking primary introductions is sensitive to the physical distance from infected groups. Each stochastic realization at a certain R_0 leads to an overall attack rate. Such stochastic-spatial models are computationally intensive if populated with millions of persons, but can evaluate a variety of interventions compared to SEIR, such as quarantines and social distancing, ring prophylaxis (where a proportion of people in a defined area around an outbreak are given antivirals), and vaccination. Such models with different assumptions and methods were constructed for Thailand (13), the U.S. and U.K. (16), and the U.S. specifically (15).

Epidemiological Modeling – Calibration to Past Pandemics: Stochastic-spatial models capture population-level effects by describing the intensity of risk at the individual level and the uncertainty in the points of origin of outbreaks as well as the distribution of epidemic-relevant factors. Such models can also be calibrated to a specific magnitude of GAR from prior pandemics. For example, a study for Southeast Asia chose to calibrate to an attack rate of 33%, based on the values from the first waves of the 1957 and 1968 pandemics (14). Additionally, the calibration can follow a relative attack rate pattern across ages (Fig. 2).

Fig. 2: Assumed pattern of attack rates across age groups at different R_0 (14)
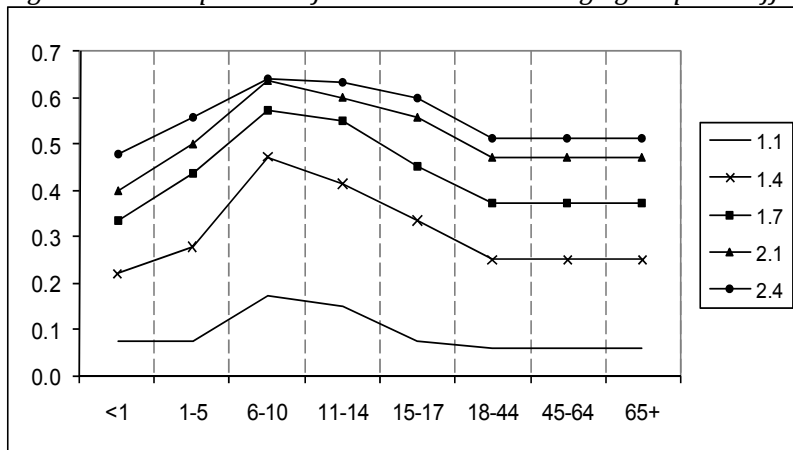
Pattern based on US data from the 1957-58 (more children infected) and 1968 (similar attack rates across ages) pandemics, mixed with seasonal influenza in SE Asia (like 1957-58 pandemic in the US).

The model produces a distribution of age-specific attack rates at different R_0. By selecting for an overall GAR of 33% and an age-specific attack rate pattern, the authors of the study determined the particular outbreak seeding that reproduces past pandemics as a base case. They tested their public health intervention strategies thereafter at this setting.

The calibrated stochastic-spatial technique captures location and demographic heterogeneity, while the full range of the stochastic results are capped at past experience. This method is good at predicting whether an epidemic that begins randomly somewhere will spread, and at evaluating which control options will be successful in limiting its spread.

Table 1 provides a comparison of a spatial-stochastic-calibrated model (14) and the age-structured SEIR model from a proposed World Bank study. The two models differ in the how they capture heterogeneity in risk of infection and in the intent. In general, some parameters are exogenously imposed constants not varied in modeling. Allowing parameters to vary implies that sensitivity to the modeling assumptions can be tested. Models can also be portable across location and demographic contexts, i.e., important variables are generated endogenously rather than assumed.

Table 1: Comparison of two model archetypes for predicting pandemic influenza transmission

Characteristic	Age-structured SEIR	Stochastic –spatial model
Source study	Part of a proposed WB study	Longini *et al.* (14)
Contact rates by age	Exogenous constants[#]	Exogenous constants
Attack rates by age	Endogenously derived	Endogenously derived[§]
Contact rates by location	Exogenous constants[#]	Exogenous constants
Attack rates by location	Endogenously derived	Endogenously derived
Whole epidemic R_0	Not modeled	Exogenously varied
Whole epidemic GAR	Endogenously derived	Endogenously derived
Raw data needs	Demographic data, seasonal influenza contact rate estimates	Demographic data, geographic data, contact rate estimates
Attack rates by policy scenario	Proposed to be modeled	Endogenously derived
Primary policy use	Describe population-level dynamics, estimate of no. of infections	Describe individual-level dynamics, test public health measures

[#] Proposed to create separate matrices of child-child, adult-child, etc., *effective* contact rates for cities by seasons.
[§] Calibrated to fit patterns of age-structured attack rates from prior pandemics at various R_0.

Modeling Public Health Interventions for Epidemic Control: A containment–focused strategy attempts to reduce of the spread of pandemic influenza within a country or internationally (13, 14). For a mitigation strategy, effectiveness refers to reducing the overall number of cases, i.e., the attack rate, and to delay and reduce the peak rate of cases per day (15, 16). A mitigation strategy lowering the overall illness attack rate unambiguously reduces hospitalizations and deaths that are related to human and

economic losses. Table 2 presents a categorized menu of control policy choices available for a potential pandemic of influenza. Italicized policies have received more research attention, or have had their effectiveness modeled.

Table 2: A taxonomy of policies for containment and mitigation of pandemic influenza

Type	Sub-Category	Individual policies (M: mitigation; C: containment)
Nonpharmaceutical	International	• International land border quarantines (C) • *Air travel restrictions & advisories* (C) • Quarantine of ships & other restrictions (C)
	National or central	• *Inter-city movement restrictions* (C, M) • Aspects of quarantine (C, M) • Public communications (C, M)
	Community or local	• *Case detection and isolation* (C, M) • *Social distancing* (C, M) • *Local quarantines* (C, M) • Public hygiene and disinfection (C, M) • Local public communication (C, M) • Personal protective equipment (C, M)
Pharmaceutical, i.e., drug-based	Vaccination	• *Targeted vaccination policies* (M) • *Broad-based vaccination policy* (C, M)
	Antiviral prophylaxis	• *Targeted (ring) prophylaxis around an outbreak* (C) • *Prophylaxis based on contact tracing* (C,M) • *Mass prophylaxis in the at-risk population* (M)

Distinctions between C and M only indicate emphasis, as containment achieves mitigation (lowers the GAR).

Recent modeling of policy interventions has followed the original proposition of a group of researchers within the MIDAS network (Models of Infectious Disease Agents Study) of a framework for control policies after the emergence of pandemic influenza. The guiding assumptions were of limited antiviral supply with no effective vaccine immediately available, and that interventions must begin rapidly at the *community level* to prevent an uncontrollable spread that overwhelms response. The proposed strategy of 'targeted layered containment' (TLC) attempts to benefit from synergies across the combined policies (28). The policies considered high priority are: targeted antiviral treatment and isolation of cases, targeted prophylaxis and quarantine of household contacts of index cases, closure of schools and keeping children at home for the duration of the policy, social distancing at the workplace (telecommuting) and in the community (cancelling public events). The assumptions on drug availability are relevant for developing countries such as India.

The results of epidemiological modeling of mixed control strategies in the TLC vein have been surveyed elsewhere (29). These studies – mostly stochastic-spatial (13, 14, 16, 30), or modified SEIR (31) – have reported threshold effects of R_0 on the efficacy of control

interventions, as well as the impact of asymptomatic infectivity (a possibility for pandemic influenza) on the effectiveness of isolation of cases and contact tracing during community-level outbreaks. The actual numerical threshold effects are model-specific and indicate how infectiousness of influenza strains can determine whether public health interventions succeed. The results on R_0 are indicative because early and pervasive public health communications can lead to a lower effective R after an outbreak than predicted by R_0, and an epidemic prevented entirely by successful containment. For example, the number of susceptibles can be reduced by forcing or urging people to limit contacts, encouraging hand-washing and other personal hygiene, or promoting the use of facemasks. The R value in the presence of interventions is usually called the 'control reproductive number' or R_c. Historically, public health communications achieved R_c much lower than R during subsequent epidemic waves when the interventions began earlier, or for similar reasons within a wave in cities and countries with later outbreaks (32, 33).

Scenario-based Policy Models of an Influenza Pandemic: Table 1 suggests that existing pandemic flu transmission model archetypes suit different policy questions. If interest is limited to impact analysis of the base case of a pandemic, then complex compartmental or stochastic-spatial transmission models may not be cost-effective. This logic drives recent non-epidemiological impact analyses based on the 1918 pandemic (11, 12, 34). The limitations of these studies were discussed above. Biological detail and the local epidemiological context are required for the scenarios to be realistic. For resource allocation decisions, the estimates should be as contextual as is feasible.

Three studies are briefly discussed that use a scenario-based model with some biological detail to inform policy decisions. A study for the U.S. used a Monte Carlo 'mathematical simulation' model to estimate the impact in the US of a pandemic of influenza and scenarios of related key interventions (35). A similar model was also used for France (36). In the U.S. study, the GAR[10] is varied in 5 percentage point increments from 15% to 35%. The total numbers of cases at a GAR are distributed using two age pattern scenarios:

Age groups	Pattern A (% of all cases)	Pattern B (% of all cases)
0-19 years old	40	46
20-64 yrs old	53.1	46.7
65+ yrs old	6.8	7.3

Based on upper and lower estimates of age-specific attack rates from 1918, 1928-29, and 1957 influenza epidemics in the U.S (1918 and 1957 were pandemics).

---

[10] They use GAR not in the sense of infections (exposures) but the number of clinical cases caused by pandemic flu per unit population during a wave.

The three age groups also differ in the proportions in each with pre-existing conditions that cause complications once an individual has influenza. These complications are health outcomes such as severe illness and death, and rates of these per age group are distributions with upper/lower limits based on statistics from past influenza epidemics in the US. The result of the study is not a single estimate of the number of clinical illnesses or a dollar figure, but a range for such variables. The authors describe their intent as "altering a number of variables and evaluating how the results affect key (policy) decisions."

In a similar study, researchers used a 'static' scenario method to test policies for pandemic influenza in the Netherlands (37). The GAR was exogenously set at 30%, and an age distribution of the cases was obtained by applying a pattern based on seasonal influenza in the Netherlands. The 'scenarios' are separate policy interventions that reduce adverse health outcomes. Compared to the U.S. study (35), the Dutch study is static since the parameters for health outcomes are point estimates from prior Dutch influenza epidemics.

Methods Used in this Study

Implications of the Survey of Methods: The discussion above has implications for a methodology given the tradeoff between model realism and parsimony, and the policy intent. The latter was noted previously as projecting epidemiologically and biologically 'realistic' by-state numbers of pandemic influenza infections under various scenarios of transmission to India, such that resource allocation decisions could be linked to the risk at both the national and provincial levels. A side benefit would be the evaluation of well-known policy interventions such as limiting air travel.

An 'explicit' stochastic-spatial pandemic influenza transmission model at the country level as recently seen (13-16) could yield a realistic value by incorporating elements of all three levels of risk differentials, as well as help evaluate nonpharmaceutical interventions for containment in the Indian context. However, this methodology is not feasible to execute at the India country level at present, even using reduced scale representation, given the high data and computational requirements. A similar logic applies to social network models for large Indian urban areas. As a metric, consider that there are at least eight Indian states with a population greater than that of Thailand, yet GIS and detailed population density data is either unavailable or of low quality. Limiting stochastic-spatial

analysis only to states identified a priori as 'high risk' for hosting an initial outbreak – i.e., the primary risk – would run the risk of omitting large resource needs in 'moderate' risk states that also have large populations. However, these analyses should be carried out to extend the results of this study. This is discussed later alongside other research priorities.

An age-structured SEIR model could be run for all major Indian cities, using separate parameter matrices for each location to capture as much variation via individual-level compartments as possible. In policy terms, this technique can help predict both natural and forced cessation of epidemics under various scenarios of epidemic initiation, and yield peak daily infection rates. However, like most SEIR models, it would suffer from a lack of spatial and population structure factors. Also, running this for a set of Indian population centers connected via air and highway transport links, across the length of an epidemic with the time step of a day, will require significant computational resources. These required resources may or may not be larger than for stochastic-spatial models. Given the unique policy insights of a modified SEIR approach, it is also a potential subject of future research.

Both metapopulation models and scenario-based non-epidemiological techniques remain feasible, and have their respective advantages and disadvantages. The disadvantages of historical, GAR-based non-epidemiological techniques were discussed earlier. A metapopulation model executed for a scale representation of India, with at least three levels of structure (local places, city, and network of cities) can reveal important issues in epidemic control. It can also be realistic for the epidemic size (GAR) if a mobility factor is introduced for symptomatic illness. In terms of policy insight, over repeated model runs it may be possible to identify target locations where a high proportion of infections occur; sensitive to the demographics, behavior patterns of groups, and the construction of the levels. Most metapopulation models also include stochasticity, usually implemented as a random seeding of initial infectives, and/or the randomly chosen distribution of the population into locations based on some limits. This randomness also implies that occasionally epidemics 'escape' outside currently infected contexts given the travel patterns of individuals, infecting unexpectedly susceptible populations, and leading to very large epidemic sizes (22). Also, resurgent epidemics can be observed in such models, i.e., epidemics which resume after a period of cessation in new infections. Without many stochastic realizations of the model (simulations), the average GAR may be very high: outside the prediction of compartmental models, the assumed relation to R_0, or the historical record for prior pandemics. The ability to test whether this is a realistic

representation of potential real-world phenomena, i.e., to run many simulations, is limited by time and resources.

The introductory section defined a framework for viewing transmission of a pandemic influenza virus to India as a set of scenarios, with subsequent internal spread leading to attack rates across states that differed on three levels of risk differentials. These scenarios of transmission could in theory be implemented with SEIR or any of the spatial-epidemiological techniques, with the model structure tuned to capture the interstate heterogeneity in risk. However, given available data and resources, and with the intent of estimating epidemic size under various conditions, this study will most benefit from a non-epidemiological, scenario-based model using weights in the vein of prior efforts (11), but with further improvements derived from scenario studies with some biological and epidemiological realism (35-37):

- *a.* Account for the variation in infection attack rates across age groups.
- *b.* Match age-structured infection attack rates pattern to recent Asian flu pandemics.
- *c.* Account for the variation in age-structured infection attack rates – both in magnitude and pattern – for different scenarios of R_0.
- *d.* Account for the variation in age-structured infection attack rates across states in India due to differing efficacy of public health systems and their governance.

In addition, there will be an attempt to capture the insights from the stochastic-spatial epidemiological modeling techniques:

- *e.* Account for the variation in the magnitude of age-structured infection attack rates across states due to different scenarios of outbreak seeding in India.
- *f.* Account for the potential difference in attack rates across locations – urban and rural – due to differences in demographics, behavior, and contact rates.
- *g.* Account for the difference across states due to the varying population densities and internal transport networks that are related to epidemic spread.

This study also executes a simplified metapopulation model for New Delhi using the software platform EpiFlex for a scale population of 300,000 (at the resolution 1:53). The results from such a structured metapopulation experiment are subject to the caveats already noted. The EpiFlex platform (25) is under further development and will in the future yield insights about rapid epidemic takeoff in, cross-immunity with seasonal influenza, and control strategies for megacities such as New Delhi. The results from the

EpiFlex simulation can be compared to the scenario results for the state of Delhi for a certain R_0. This is discussed further below in the presentation of the EpiFlex model.

Disease Entry Scenarios and State Risk Differentials: The two specific scenarios for disease transmission to India were discussed in the introduction, along with one worst-case scenario where both types of transmission occur (see Table 3 below). In the quantitative analysis of the scenarios, the highly simplifying assumption is made that the number of infectives arriving per day in the Air scenario (initial outbreaks in India due to transmission via air) is the same as that in the land-sea scenario (initial outbreaks due to transmission via individuals entering at seaports and land border crossings). Given this assumption, the difference at the countrywide level of total infections results across the two scenarios derives from the three levels of risk differentials across the states, beginning with the primary and most important differential based on whether a state holds an international airport (air-entry scenario), or a seaport or land border crossing (land-sea entry scenario).

In actuality, the inherent assumption of a land-sea entry scenario is that Indian authorities successfully prevented the importation of infection by air. In this scenario, when pandemic influenza prevalence in neighboring and maritime trading partner countries reaches a high enough level, then it is impossible to prevent the entry of infected individuals across the land and sea borders. However, the daily volume of such traffic – and hence the volume of potential infectives – is low compared to volume of daily air passenger traffic entering India. Therefore, even prior to applying the three levels of risk differentials, the potential for outbreaks in this scenario, and hence the potential state-level attack rates, could be lower. With better data on land border crossings and seaport passenger/crew traffic at the state level, it may be possible in the future to account for this issue.

Table 3: Scenario-based prediction of pandemic influenza in India: assumptions

Transmission Scenario (name)	Assumed internal spread dynamic
1. Fixed GAR approach, no state weights (**Base Case**)	Uniform, rapid and simultaneous spread across states
2. International air travel based entry of disease (**Air Scenario**)	*First stage*: In and around major cities with intl. airports *Second stage*: Other cities with a major domestic airport *Third stage*: Small towns, rural areas in states with stages 1 & 2
3. Land border & seaport based entry of disease (**Land-Sea Scenario**)	*First stage*: Towns in Northeast states, eastern seaport cities *Second stage*: Western seaport cities, rural areas in Northeast states *Third stage*: Other cities with a major domestic airport
4. International air, land border, and seaport based entry of disease (**Worst Case Scenario**)	Sum of stages above for 2 and 3.

Table 3 presents the assumptions for each scenario vis-à-vis the likely dynamic for internal spread of pandemic influenza, as per the location and size of outbreaks in each scenario's type of disease transmission to India. Overall, these assumptions guide the choice of various variables that are used to combine the primary and secondary risk differentials in the form of a single state-level *index of geographic risk* (see Table 4 below). The primary risk differential across the states in the *Air scenario* derives from the presence or absence of international airports, which are located in major metropolises (New Delhi, Mumbai, Bangalore, Kolkata, and Chennai). The primary risk for the *land-sea scenario* derives from the presence of international sea ports, or land border crossing points on the territory of the state. Based on the data on seaport shipping and crew traffic (across ports on India's eastern vs. western seaboard), and unofficial data on illegal border crossings on the eastern borders, the stages for the land-sea scenario in Table 3 reflect the risk differential and the timeline across states.

Across scenarios, given the primary risk, the secondary risk differential reflects the dynamics of how outbreaks spread and become epidemics, involving the density of urban and rural population, and state transport networks (roads and domestic air travel). It is assumed that the timeline of stages for each scenario in Table 3 includes enough warning by the third stage for central authorities to close domestic air and rail traffic and/or impose quarantines around areas with confirmed fatalities in the earlier stages. Therefore, areas not mentioned in Table 3 have the lowest risk. A separate geographic risk weight is calculated for rural and urban populations. Therefore, there are two schemes of state-level weights, rural and urban, for each scenario 2-4 in Table 3.

Table 4: Composition of the index of geographical risk

Scenario	Index of geographical risk: factors by state				
	Primary risk		Secondary risk		
Air : urban	Avg. monthly intl. air passenger movements		Pop. density (urban)	Avg. monthly domestic air passenger movements		Highway kilometers
: rural	Intl. airport in state; In-state road density		Pop. density (rural)	(domestic air) – do –	Highway kilometers	Avg. % of villages connected by road
Land & Sea	Intl. border factor	Port traffic	As above, applied as per urban or rural populations.		

Prompt and efficient community control interventions could reduce infections. This is the tertiary level of risk differential discussed earlier. The ability to take such action would vary across state governments. These weights are fixed across both rural and urban areas. This is a simplification since health system preparedness may is better in urban areas

(though coordination problems are greater). The variables used to create this *index of state health system preparedness* track both *resources* and *efficacy* (see Table 5 below). Only variables with data for all states could be used. The effectiveness of interventions may also depend on population characteristics (literacy, density, etc) that are difficult to account for.

Table 5: Composition of the index of state health system preparedness

Index of health system preparedness (tertiary risk differential)	
Resources	Per capita public health expenditure; Numbers of district hospitals, sub divisional hospitals, health sub-centers, primary health centers, & community health centers
Efficacy	Infant mortality rate

Details of the variables used, data sources, and calculations for both indices above are provided in the supplementary materials appendix. The numerical score of a state on an index determines whether it is assigned a weight between 0 and 1, from the set: lowest, low, medium, and high. The baseline assignments of values to the weight labels are as follows:

Weight labels	Index of geographic risk	Index of state health system preparedness
Lowest	0.25	0.95
Low	0.5	0.8
Medium	0.75	0.66
High	1	0.5

The values above indicate that a state with high geographic risk suffers 100% of the attack rates shown in Table 4 below (at a certain R_0). The state with a high weight related to the index of health system preparedness reduces geographical risk-weighted attack rates by 50%. In contrast a state with a weight of 'poor' on the same index reduces it by just 5%. Sensitivity analyses are conducted on the decimal values of the weights and are discussed later. Equations 1.1-1.2 summarize how the weights are applied.

(0.1) $SI_i = (w_{irural} \sum_{j=1}^{8} a_j.p_{ijrural} + w_{iurban} \sum_{j=1}^{8} a_j.p_{ijurban})$

(0.2) $SI_i = h_i(w_{irural} \sum_{j=1}^{8} a_j.p_{ijrural} + w_{iurban} \sum_{j=1}^{8} a_j.p_{ijurban})$

Where SI_i = Total pandemic influenza infections in state i during a single wave

w_{ix} = Weight for index of geographic risk for location type x (urban/rural),

    for state i ($0 < w_{ix} < 1$): primary and secondary risk differential

h_i = Weight for health policy preparedness (tertiary risk) differential for state i ($0 < h_i < 1$)

a_j = Attack rate at the pre-defined level of R_0 for age group j of eight (fixed across states)

p_{ijx} = Population in age group j in location type x (urban/rural) for state i

Improvements in the scenario-based model include that age structure and historical patterns be accounted for in the attack rate of pandemic influenza. The pattern in Fig. 2 above is laid out in Table 5 and is based on the analysis of the two most recent influenza A pandemics and inter-pandemic influenza in the U.S. and Southeast Asia (14). This pattern will be applied to the demographic data on Indians in different age groups in each state. The demographic data is available and tabulated by rural and urban splits, which means that the final output will be infections by age, location, and state, for different scenarios of R_0.

Table 5: Gross infection attack rates across age groups at different R_0 (14)   *Also see Fig. 2*

R_0	<1 yr	1-5 yr	6-10 yr	11-14 yr	15-17 yr	18-44 yr	45-64 yr	65+ yr
1.1	0.07	0.07	0.17	0.15	0.07	0.06	0.06	0.06
1.4	0.22	0.28	0.47	0.41	0.34	0.25	0.25	0.25
1.7	0.34	0.44	0.57	0.55	0.45	0.37	0.37	0.37
2.1	0.40	0.50	0.63	0.60	0.56	0.47	0.47	0.47
2.4	0.48	0.56	0.64	0.64	0.60	0.51	0.51	0.51

Fig. 3 Index of geographical risk weights, Air scenario, for only the urban populations of states
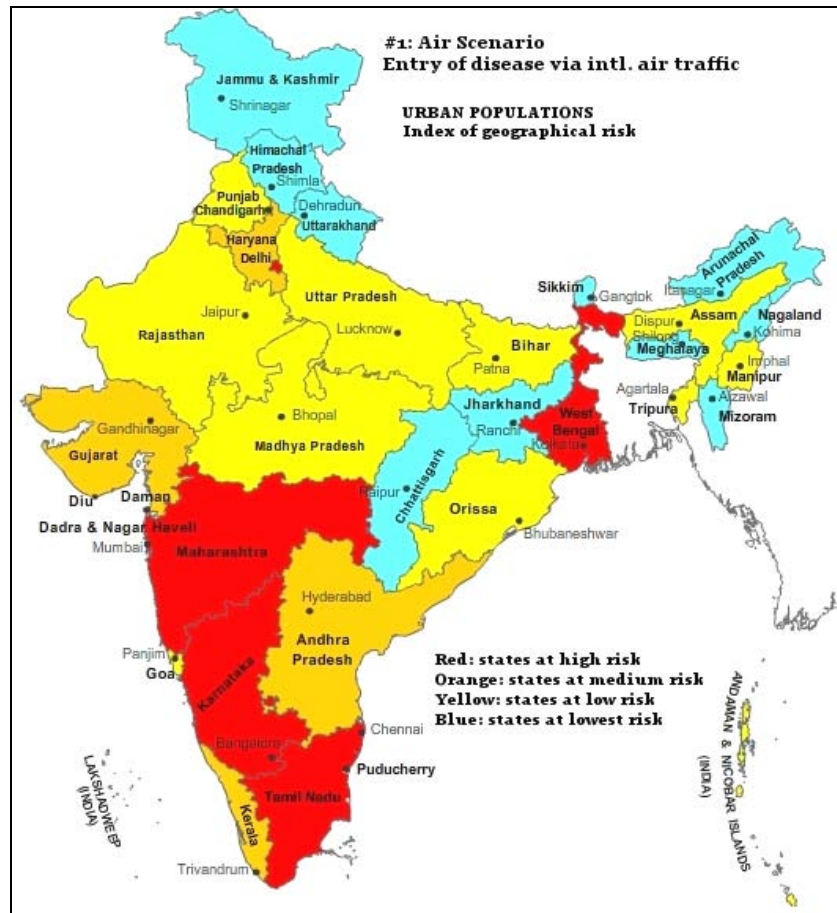
Also see Table 3.1 in the supporting materials.

Fig. 3 above shows the results of applying both primary and secondary levels of risk differentials to the urban populations of the thirty-two states of India under the Air scenario. The result is four levels of weights based on the underlying index of geographic risk weights applied to states in the form of a four-color scheme. Similar maps for urban populations under the Land & Sea and the worst-case scenario, along with tables of the weights used for all three scenarios and for both locations, are available in the appendix. Additionally, the weights based on the index of health system preparedness are listed there.

Structured Metapopulation Modeling in EpiFlex: An experiment is conducted in this study using the EpiFlex software where an epidemic is caused by the forced seeding of pandemic influenza in a scale model of New Delhi. The scale population of 300,000 compares to the estimated 2007 population of 16 million in Delhi state at a 1:53 ratio. The reduced scale was necessary for computational reasons. The EpiFlex software is described elsewhere (25). In brief, EpiFlex is a software environment in which infectious disease events can be simulated at the level of the individual and where attention is paid to demographics and subgroups and the contacts between these groups. In a comparison experiment using a seasonal influenza model, EpiFlex results reproduced the pattern seen in actual WHO/NREVSS surveillance data (25).

In the language of metapopulation models introduced before, the software allows three levels of structure: disease *hosts* who are a part of specific demographic groups (with behaviors such as movement cycles specific to their group); *locations* which are temporary containers for hosts from different groups as they go through their movement cycles; and *areas* which are containers for various configurations of locations. In the executed model, there are 14 groups, 12 locations, and only one area (New Delhi city). Locations can be repeated within an area using *N* concurrently active, identical cells which have user-defined average numbers of hosts in them at a point of time.

The disease model for pandemic influenza as executed for this study in EpiFlex includes the three conventional states from compartmental models: susceptible, infected, and removed (immune or dead). However, the infected period is split into three separate stages with increasing infectivity given effective contacts: incubating (no or very low infectivity), prodromal (onset of disease, early and potentially hard to diagnose symptoms and mild infectivity), and manifestation (pronounced symptoms, full infectivity). Hosts can be infected using a variety of contacts with influenza virus carried as aerosol, fomites, or

actual mucosal secretions on surfaces. The movements of hosts are circumscribed once they enter into symptomatic disease stages. Case fatality rate for pandemic influenza disease is fixed in the model, and using literature review of various sources, this was set at 1.3% - a moderately virulent form of a 1918-like virus[11]. Various seed vectors into the defined area were tried, yielding different results which are discussed later. A realistic vector tied to limited introduction of infectives such as the Air scenario, sees a fixed number of infections forced every five hours for fifteen hours in New Delhi, before authorities close the avenue of importation of disease, e.g., by suspending incoming flights from certain cities.

The model is run by first allocation hosts to their demographic group based on the estimated population percentages of those groups in the target city. Stochasticity is introduced by assigning contacts to hosts at locations using a Monte Carlo algorithm that is driven by a user-defined contact skew curve – highly skewed curves imply that a few highly active actors provide most of the contacts (i.e., 'super-spreaders'). Based on this author's experience of New Delhi social contact places and contexts, unique contact skews were applied to the 12 location types used in the model. The description of the levels and the other aspects of the implemented model in EpiFlex are provided in the supplementary materials. Results of the EpiFlex model runs – including a public health response as expected in a city like New Delhi – are discussed further below. The value from this run can be compared to the value from the scenario run for Delhi state, at various levels of R_0.

Results

Choice of a Reference Level of R_0: As the R_0 values in Table 5 were increased, attack rates across age groups increased in a non-linear fashion, especially with a big jump between 1.1 and 1.4. The relationship in Table 5 between increasing infectiousness as proxied by R_0 and attack rates is an artificial one; the former is estimated from the epidemiological data which may or may not reflect the entire epidemic, and attack rates can be high for reasons other than the virus's biological infectivity (for example, due to cramped barracks during 1918). In Table 5, the increasing R_0 leads to increasing epidemic size and a change in its distribution across age groups. The attack rates are especially high for younger

---

[11] Choice of the case fatality rate determines the overall mortality; but in this model, it does not determine transmission. Instead morbidity – modeled as incapacitation during the symptomatic stage – affects the movements of hosts, therefore the transmission process, and hence the attack rate.

age groups between 6-17 years, and this pattern becomes more pronounced as R_0 increases. This is significant for developing countries like India where 40% of the population is below the age of 20 (2006 population estimates). Given these points, the choice of R_0 is crucial to evaluate the results of the weighted scenarios described previously. Table 6 below reports the estimated R_0 values for various outbreaks of pandemic influenza in the past as well as for SARS.

Table 6: Estimates of R_0 for past outbreaks (citation/source)

Pandemic influenza, 1918, New Zealand	$1.3 < R_0 < 3.1$ (38)
Pandemic influenza, 1918, UK (spring wave)	$1.7 < R_0 < 2$ (16)
Pandemic influenza, 1918, Geneva (spring wave)	$1.45 < R_0 < 1.53$ (18)
Pandemic influenza, 1918, Geneva (fall wave)	$3.57 < R_0 < 3.93$ (18)
Pandemic influenza, 1968-69, Hong Kong	1.89 (39)
SARS 2003 Hong Kong	$2 < R_0 < 4$ (30)

The results from Table 6 indicate that the estimated R_0 in past influenza-A pandemics has been well above 1.4. In moving above $R_0=1.1$ in Table 5, the attack rates increase by three- or four-fold, depending on the age group. Therefore, there is a very large impact on the epidemic size, as can be seen in the results of the scenario analysis with the index of geographical risk weights imposed (Table 7).

While it is unknown what the initial R_0 will be for any pandemic derived from the current H5N1 influenza-A subtype, there is the force of history to suggest the range will be between 1.1 and 4. Other studies have focused on $R_0=1.4$ as the reference level for further analysis (14). However, the excess death estimates in Indian provinces during the 1918 pandemic (20), given the population at the time suggest a high attack rate at case fatality rates similar to those recorded elsewhere. This in turn indicates a higher R_0. Alternatively, case fatality rates in India were higher than elsewhere. Given that neither is the exact cause of the 1918 pattern in India knowable, nor is there any guarantee of the pattern repeating in a future pandemic in the country, the choice of R_0 for analysis has to be based on the current literature. The choice of $R_0=1.4$ as the reference level in this study is an inherently conservative one, viewing this value as an optimistic figure. This value represents a compromise: while better viral prophylaxis is available even in developing countries compared to the era of past influenza pandemics (hence downward pressure on R_0), but population density and potentially contact rates have increased, and cross-immunity from

prior pandemics[12] (40) or from seasonal influenza may be a less significant factor (hence upward pressure). The choice of a reference R_0 is significant. At R_0=1.4 the epidemic size in terms of total infections in India across scenarios is roughly 3.6 times that at R_0=1.1.

Table 7: Potential pandemic influenza infections in India by scenario and R_0 (millions)

R_0	Base/Rapid	Air	Land & Sea	Worst Case
1.1	92	61	34	62
1.4	335	222	124	227
1.7	470	312	174	319
2.1	567	377	211	386
2.4	613	407	228	417

Results of Scenarios with Geographic Risk Weights: The 'Base/Rapid' scenario is presented once only for reference in Table 7. It is the result if the attack rates at various R_0 as in Table 5 are applied without weights to the demographics of the states. This scenario can be understood as the implausible situation that the pandemic spreads rapidly and equally across rural and urban areas of all states, infecting individuals as per the pattern seen in Table 5, which is itself a schema constructed from the U.S. record in the 1957-58 and 1968-69 pandemics, as well as seasonal influenza in Southeast Asia (14). Compared to the 'Base/Rapid' scenario, the Air scenario epidemic size is 32% lower on average across R_0 values and that for Land & Sea scenario is 63% lower.

The Air and Worst Case scenarios are similar in epidemic size, as the net effect of adding 'Land & Sea' transmission to India to the Air scenario is the change in the geographic risk weights on a few states, given that the Worst Case is constructed as the sum of the two other scenarios and hence has the higher (worse) weight of the two.

The values in Table 7 can be compared to non-epidemiological scenarios for India as in some recent studies (11), where a flat GAR of 30% was used, implying 316 million infections. From Table 5, at R_0=1.4, almost one-third of India's population, i.e., in the age groups 6-10, 11-14, and 15-17 years (2006 population estimates), have unweighted attack rates *above* 30%. However, in this study the Air scenario yields 222 million infections, or a GAR of 21%, and the Land & Sea scenario at 124 million yields a GAR of 12%. This suggests that the geographic risk weighting scheme matters.

---

[12] In 1968, where the influenza A subtype causing the pandemic was H3N2, the virus differed from the 1957 pandemic in its HA (hemagglutinin) antigen but shared the same NA (neuraminidase) antigen, offering some variable protection to specific populations with prior N2 immunity (37). While some N1 (the same NA antigen type as in H5N1) immunity may be present from previously circulating seasonal influenza virus subtypes, this may be over-represented in older cohorts.

Figures 4-a and 4-b display the distribution of the potential infections by age and location. While urban areas of India account for 29% of the estimated 2006 population, they involve about 32% of the infections under the Air scenario and 42% under the Land & Sea scenario at R_0=1.4. The infections by location for each scenario are also higher as a percentage of the urban population than as a percentage of the rural population.

At R_0=1.4, the age group 0-19 years (i.e., children) accounts for 51.5% of the infections, the age group 20-64 (i.e., the working age population) accounts for 44%, with the rest accounted for by the elderly. The pattern changes slightly with a higher R_0, to 48.7%, 46.7%, and 4.6% respectively. The percentages across the age groups are similar across the two discrete scenarios: Air vs. Land & Sea. In comparison to their population shares, the 0-19 group is disproportionately affected, as suggested by Fig. 2. This disproportionality eases slightly against the 20-64 age group as R_0 rises.

Fig. 4 a-b: Distribution of potential pandemic influenza infections by location and age
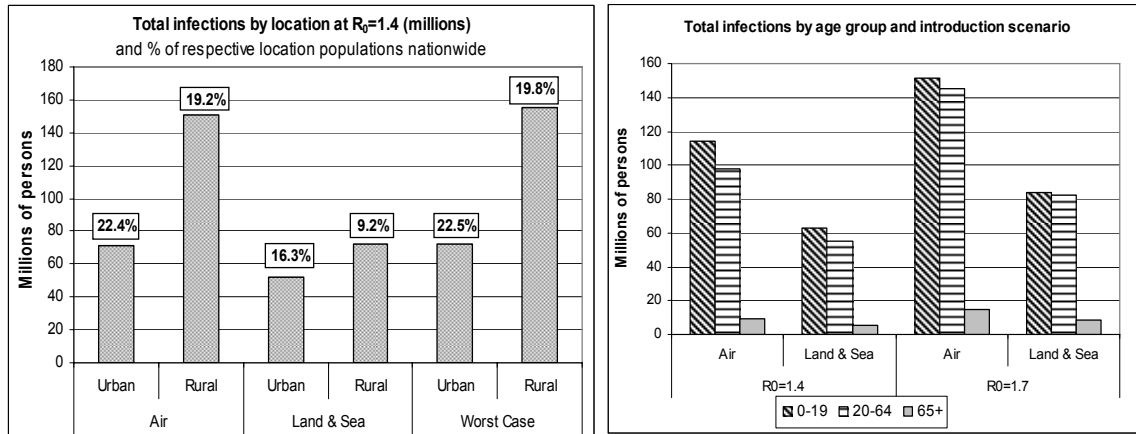
Results with Geographic Risk Weights and Asymmetric R_0 by Location: The variation in the two sets of geographic risk weights, one each for urban and rural areas by state, captured the difference at the level of population density, transport networks, and proximity to first outbreaks and movement of infectives across the two locations. These considerations usually favor a lower weight for rural populations (see supplementary materials). However, these considerations at the geographic or population level do not include the differences at the level of the average rural person vs. the average urban person in India. Given differences at such an individual level, a case can be made that rural areas will face a much lower reproductive number than the urban areas. Another way to

understand this is that the 'effective reproductive number' or R at the start of rural epidemics will be lower in rural areas at any overall *a priori* basic reproductive number value (faced in full by urban areas).

Though not all rural areas in India are comparable, in general the following argument can be made. As can be recalled, the determinants of the effective reproductive number include the number of effective contacts made by a representative individual in the population. A rural individual meets fewer persons as a part of their daily routine and also does not meet such contacts frequently in a day given their movement patterns. Also, rural individuals are less exposed to other risks of transmission, such as infected mucosa from surfaces, or virus-bearing aerosols in crowded, confined spaces such as elevators, theatres, and public transport. The argument that rural individuals might be more susceptible to infection – as distinct from the susceptibility to clinical complications post-infection – because of poorer health is not strong, given that a pandemic virus would attack even healthy individuals. However, cross-immunity due to prior seasonal influenza epidemics, if present, could favor urban dwellers more than rural, given that the former have greater seasonal influenza exposure (almost year-round in the tropics).

The exact difference between the averaged effective R in rural areas vs. that in the urban areas is difficult to fix in analysis, given the uncertainty over the importance of the factors listed just above. However, if the interest is to analyze the broad impact of this distinction for overall epidemic size in India, and to utilize the available mapping of R_0 to attack rates by age as in Table 5, then it can be assumed that if urban areas face the attack rates specific to a particular R_0 (e.g., 1.4), rural areas face those relevant to the R_0 level just below it (i.e., 1.1). Table 8 below reports the results with the three scenarios, using the index of geographical risk weights and the asymmetric R_0 patterns by location.

Table 8: Potential pandemic influenza infections in India by scenario, asymmetric R_0 (millions)

Rural R_0 / Urban R_0	Air		Land & Sea		Worst Case	
	Same	Lower	Same	Lower	Same	Lower
1.1	62	N/A	34	N/A	62	N/A
1.4	222	*113*	124	*72*	227	*114*
1.7	311	*251*	174	*145*	319	*256*
2.1	377	*333*	211	*190*	386	*340*
2.4	407	*386*	228	*218*	417	*396*

At urban R_0=1.4 and rural R_0=1.1, the epidemic size for the Air scenario is 49% that using the symmetric R_0, and similarly 42% for the Land & Sea scenario. The reduction in the

epidemic size at the national level using the asymmetric R_0 values declines as R_0 in the urban areas rises, simply because the difference in the attack rates as seen in Fig. 2/Table 5 across R_0 values also declines with rising R_0. In addition, at urban R_0=1.4, the share of infections in children (0-19 years) also rises compared to the symmetric case (for details, see the supplementary materials appendix). The implication of the difference in individual-level determinants of the reproductive number of pandemic influenza epidemics across location type is strongest for *moderate* infectiousness of the virus.

Results with Geographic and Health System Preparedness Weights: Table 9 reports the results of applying the health system preparedness weights (fixed at the state level across urban and rural locations). Since the weights are below 1, the result is the reduction of the attack rates commensurate with the state's score on the index of health system preparedness. Summing the weighted potential infections across states, Table 9 shows the totals at different values of R_0 at the level of the country.

Table 9: Potential pandemic influenza infections in India by R_0 and type of weights (millions)

	Air		Land & Sea	
Weights R_0*	Geographic Risk only	Geographic Risk & Health Sys.	Geographic Risk only	Geographic Risk & Health Sys.
1.1	62	*39*	34	*22*
1.4	222	*142*	124	*80*
1.7	311	*199*	174	*112*
2.1	377	*240*	211	*136*
2.4	407	*260*	228	*147*

* Symmetric R_0 across rural and urban locations

Across the two discrete scenarios in Table 9 – i.e., omitting the Worst Case scenario – the total infections with the health system preparedness weights added are lower by 35.5%-36% across the various values of R_0. This more than one-third reduction is the proxy for the mitigation in attack rates achieved as states initiate nonpharmaceutical interventions such as local quarantines, health and hygiene measures in public areas, social distancing in cities, as well as public health communications. The ability to initiate and sustain such interventions during the epidemic wave in order to bring about a lower level of the control reproductive number R_c, will vary across states. Of the 32 Indian states and union territories included in this study, 10 had a weight of 'Lowest' (5% reduction in attack rates at any R_0), 9 had a weight of 'Low' (15% reduction), 9 were 'Medium' (33% reduction), and only 4 were 'High' (50% reduction). Sensitivity analyses were conducted on the percentage reduction assigned to the various weight labels that are based on the state's

score on the index of health system preparedness. These analyses and the individual weights by state are presented in the supplementary materials.

Figures 5-a and 5-b summarize the analysis so far from applying the index of geographic risk weights at a symmetric R_0 across urban/rural areas, then at asymmetric R_0, and then with the index of health system preparedness weights across both symmetric and asymmetric R_0 situations. At the reference R_0 level of 1.4 (i.e., the asymmetric 1.4/1.1 case), given the most optimistic analysis – applying health system preparedness weights *as well as* asymmetric R_0 – there is a reduction of 68% in the Air scenario (Fig. 5-a), and a reduction of 63.7% for the Land & Sea scenario (Fig. 5-b). At 45 million infections, the latter value for the Land & Sea scenario is the lowest forecast at R_0=1.4 for potential pandemic influenza epidemic size in India. This forecast is 14% of the value from applying a flat 30% GAR. The reductions between the 'geographic risk weights/symmetric R_0' patterns to the 'geographic plus health system weights/asymmetric R_0' are lower at higher levels of R_0. For example, at R_0=2.4, the reduction for the Air scenario is 39.4%, and 38.5% for the Land & Sea scenario.

Fig. 5 a-b: Potential infections by scenario, health weights, and symmetric/asymmetric R_0
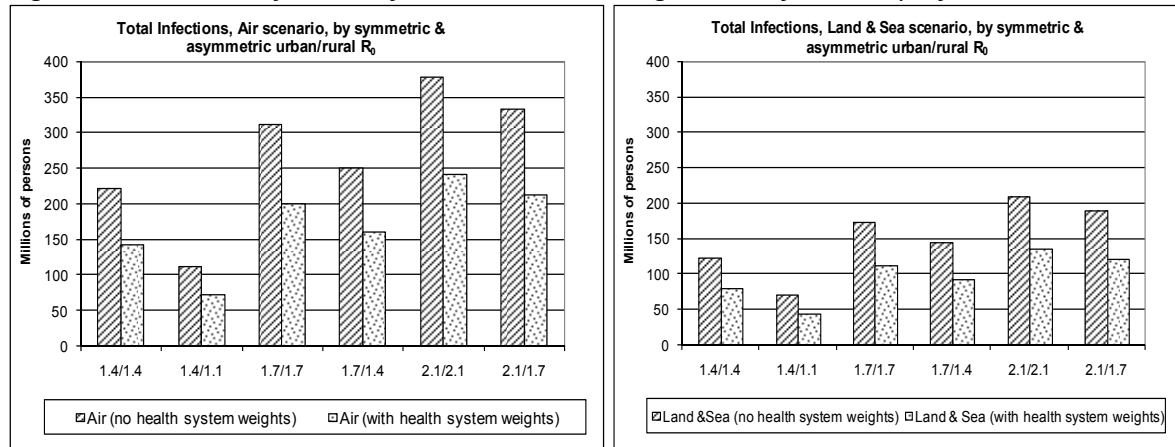
Sensitivity Analyses: The results of varying the underlying weights used in the analyses above can be seen in lieu of probabilistic or confidence-limits analysis for the estimates of potential infection. In Figs. 6-a and 6-b, the sensitivity results are reported for the total national infections under the two discrete scenarios of entry, from varying the decimal values for the geographic risk weight labels 'Lowest' and 'Medium' (holding the values for the other risk labels constant). A maximum of two weights can be varied at a time. The values for these risk labels were fixed for the baseline analysis at 0.25 and 0.75;

implying that states with the 'Lowest' risk label would face 25% of the attack rates at the reference R_0 level of 1.4, and similarly, 75% for states with the 'Medium' label. In Figs. 6-a and 6-b the pairs of weights on the X-axis range from the *optimistic* 0.15/0.66 to the *pessimistic* 0.4/0.9. The charts imply that the range of estimated infections under the Air scenario for symmetric rural/urban R_0=1.4 is 207-246 million (baseline value: 222 million), and the same for the Land & Sea scenario is 99-162 million (baseline: 124 million). The Land & Sea scenario is more sensitive to variation in these weights, given that more states have the geographic risk labels 'Lowest' and 'Medium' than in the Air scenario.

Fig. 6 a-b: Sensitivity to geographic risk weights: Air, Land & Sea scenario estimates
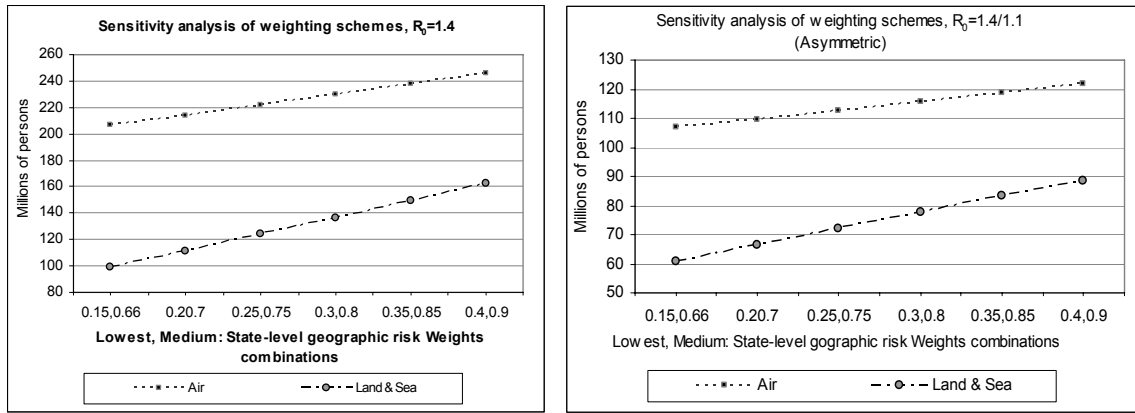
The opposite is true for the sensitivity of scenario results to weights attached to the index of health system preparedness. The baseline assignment of fractional values to the labels 'Lowest' and 'Medium' was 0.95 and 0.66, implying states with the former label reduced attack rates due to public health interventions by 5% and the latter labeled states reduced these by 34%. The results under the Air scenario at symmetric R_0=1.4 (baseline: 142 million), are more sensitive to varying the health system weights – holding geographic risk weights constant – ranging from 114-151 million across the optimistic (20% and 60% reductions respectively) to the pessimistic (1% and 25% reductions). In the same context, the Land & Sea scenario varies much less: from 75-82 million (baseline: 80 million). The reason can be traced to the Air scenario having more states with large infection estimates.

Fig. 7: Sensitivity to health system weights, ceteris paribus: Air, Land & Sea scenario estimates
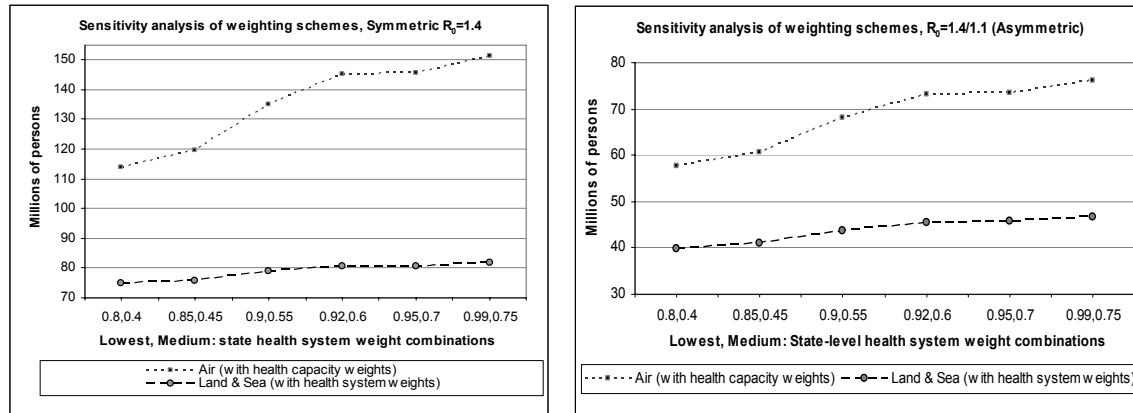
Results of the New Delhi Metapopulation Experiment in EpiFlex: Eighteen model configurations following the schema described in the *methods* section were run on the EpiFlex software platform. The GAR in New Delhi with initial metapopulation forms that mimicked U.S. cities was in the range 98-99%. These model runs resulted in further refinement of the structure of the metapopulation best able to track the urban demographic and behavioral map of New Delhi. The final model configuration with a conservative rate of contacts of various types at city locations was run with two patterns of probabilities of transmitting infection per contact, by type of contact (Table 10). The probabilities specified are purposively below the fixed 10% probability of transmission per effective contact used in recent studies (14). The rationale and other details are given in the appendix.

Table 10: Contact infectivity patterns for a New Delhi population of 300,000

Contact type	Probability of transmitting infection per contact (%)	
	Pattern I	Pattern II
Close airborne (aerosol)	4.5%	1.2%
Casual airborne (aerosol)	2.5%	1%
Surface-to-hand (mucosa)	1.5%	1.2%
Food-to-hand (mucosa)	1.2%	1.2%

The epidemic curves – plot of persons in various disease states over time in days – with each pattern above are displayed in Figures 8 and 9. The epidemic curves for the two patterns are strikingly different even if the only modeled variation is in the contact infectivity per effective interaction. The case fatality rate is fixed at 1.3%. The public health response in both runs is modeled as a decline in contact infectivity (e.g., due to facemask use, personal hygiene) rather than in the rate of contacts at various city locations (as expected with the nonpharmaceutical interventions in Table 2). Other, spatially defined

interventions such as ring prophylaxis and area quarantine schemes, if modeled to realistic standards, would be a complex addition to structured metapopulation modeling. These may be added to EpiFlex in the future (25). Given the results, the choice between these two types of nonpharmaceutical policies is taken up in the *Discussion* section.

Fig. 8: Epiflex pandemic influenza epidemic curve in New Delhi, contact infectivity Pattern I
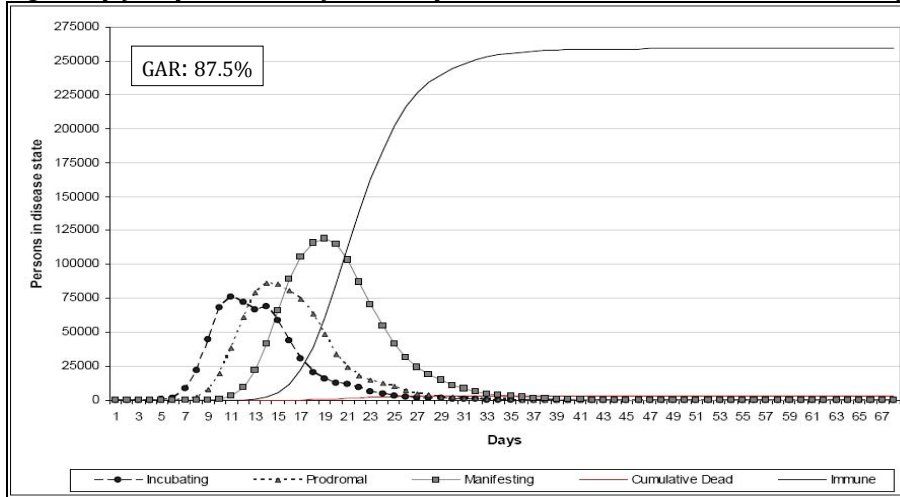
Fig. 9: Epiflex pandemic influenza epidemic curve in New Delhi, contact infectivity Pattern II
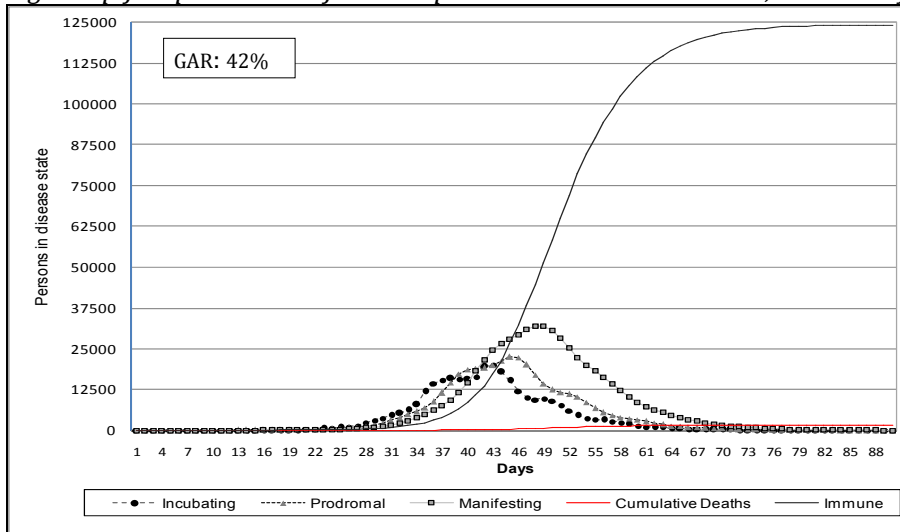
In Fig. 9, the lower infectivity pattern reflects in a curve which is compressed and shifted to the right, implying substantially fewer clinical cases as well as increased lead time before the peak of the epidemic hits the simulated city. The actual size of the epidemic – i.e., the gross attack rate – should be taken as indicative given that no actual field study of contact rates at locations was undertaken for the metapopulation study. Given the caveat,

the GAR is more than halved from Pattern I to II, yielding a 42% figure in the latter for individuals in either immune or dead (from disease) states at the end of the main epidemic.

This result is high compared to the GAR values from the scenario-based analysis. Table 11 shows the potential infections under the two discrete scenarios, evaluated as the GAR over the entire epidemic wave, i.e., as a fraction of the total urban and rural population of 16 million of Delhi state in 2006. Reasons for the discrepancy between these and the EpiFlex results, and the implications therein are presented in the next section, *Discussion*.

Table 11: GAR from weighted scenario-based analysis for Delhi state, at total pop. of 16 million

	Air scenario		Land & Sea scenario	
Weights used	Sym. R_0	Asym. R_0	Sym. R_0	Asym. R_0
Geographic risk only	29.7%	28.5%	14.4%	14.1%
Geographic risk & health system	28.2%	27.1%	13.7%	13.4%

Discussion

As referred in the *Introduction*, this discussion is in two domains. The first domain concerns benefits of costly prevention *interventions* against the *a priori* 'low probability, high impact' public health event such as a countrywide pandemic influenza epidemic in India when no related international outbreak in humans has yet been reported. The second, discussed further below using the results of both the scenario-based and EpiFlex modeling, concerns *research* priorities to understand the spread of and interventions in the control of local epidemics in India when pandemic influenza outbreaks have already occurred globally (i.e., in the WHO pandemic alert phase five). In this *ex post* context, a pandemic influenza epidemic in India is a 'high probability, high impact' public health event.

Implications for Public Health Policy: Towards the determination of *benefits of costly prevention measures* such as closure of international borders in a world without the announcement of a WHO pandemic alert phase five (large clusters of localized human-to-human transmission, as a precursor of a pandemic event or phase six), it is necessary to understand the costs in the no intervention future. These costs relate to health and economic shocks (where monetary values are not additionally imputed for morbidity and mortality). All health and economic costs of an epidemic fundamentally relate to how many

of those infected in an epidemic will be moderately, seriously, or fatally ill. The determinants of such morbidity and mortality estimates include the number of infections (related to the base infectiousness or R_0 of the viral strain and all the factors qualifying human-to-human transmission as outlined previously) and the base virulence of the viral strain (qualified by individual-level and local health system-level factors).

Using scenarios of fixed attack rates (epidemic size) and virulence, health and economic shocks on the economy have been calculated for various countries and regions, based on macroeconomic models of various types (11, 12, 34). Table 12 compares possible epidemic size realizations under the two discrete scenarios of disease transmission to India used in this study, assuming that R_0 will be 1.4. Realizations under these two scenarios vary from 13% (Land & Sea, Very Optimistic) of the estimate using a fixed 30% epidemic size (316 million), to 70% of that value (Air, Pessimistic).

Table 12: Pandemic influenza epidemic sizes in India by scenarios at base R_0=1.4

	Effective R scheme and weights mix			
Scenario type	Very Optimistic**	Optimistic#	Moderate§	Pessimistic ξ
Fixed 30% GAR nationwide*	316 million (11)			
Air Scenario	58 million	71 million	107 million	222 million
Land & Sea Scenario	40 million	45 million	61 million	124 million

* No R_0 specified. Population for India suggested from mortality/morbidity rates & estimates: 1.054 billion.
** Asymmetric urban/rural R_0, use of optimistic values for state health system preparedness weights
# Asymmetric urban/rural R_0, use of state health system preparedness weights (at baseline values)
§ Asymmetric urban/ rural R_0, use of optimistic values of geographic risk weights
ξ Symmetric urban/rural R_0, baseline values of all weights

The health and economic shock estimates based on applying a fixed, non-age-structured GAR to the Indian population as in two recent studies (11, 12) can be rescaled by comparing the inherent epidemic size of these estimates to the epidemic sizes under the various scenarios under Air entry and Land & Sea entry (Table 13). The economic shock estimates selected to be rescaled in Table 13 are at two levels of virulence. Type I shocks (11) are at moderate, and Type II shocks (12) at severe case fatality rates. The two types are calculated using different macroeconomic models.

The shocks to economic output in Table 13 include in the very short run those from absenteeism due to illness or care-giving activity of workers, and in the short run due to the loss of skilled workers from mortality which reduces output till replacements are trained or hired (11, 12, 34). Also included is the decline in demand as individuals self-protect by

curtailing movements, or do so due to quarantines and officially mandated social distancing; which severely affects the economy in the short run (11, 12, 34).

Not included in Table 13 are supply effects that persist into the quarter after the pandemic wave, as firms reestablish operations and inventories (11). The rationale for this is that at the aggregate, annual level, the upsurge in consumption after the epidemic would make up for the residual supply shocks. Potential financial shocks due to the movement of portfolio capital in response to divergent risk profiles of countries affected by a global pandemic (11) are not included.

Table 13: Scaled mortality and macroeconomic shocks in India, based on scenarios in Table 12

Epidemic size (Source)	Mortality[1] ('000s)			Economic shocks, US$ millions (GDP%)[2]			
				Labor Shock		Demand shock	
	Low	Mod.	Severe	Type-I[3]	Type-II[4]	Type-I[5]	Type-II[6]
McKibbin *et al.* (11)	74	737	5,267	-$3,916 (-0.55%)	-	-$997 (-0.14%)	-
The World Bank (12)	86	860	6,145	-	-$9,969 (-1.4%)	-	-$15,665 (-2.2%)
Rescaled: Air scenario							
Very optimistic	14	135	967	-$719 (-0.1%)	-$1,567 (-0.22%)	-$183 (-0.03%)	-$2,463 (-0.35%)
Optimistic	17	166	1,183	-$880 (-0.12%)	-$1,919 (-0.27%)	-$224 (-0.03%)	-$3,015 (-0.42%)
Moderate	25	250	1,783	-$1,326 (-0.19%)	-$2,891 (-0.41%)	-$338 (-0.05%)	-$4,544 (-0.64%)
Pessimistic	52	518	3,700	-$2,751 (-0.39%)	-$7,003 (-0.84%)	-$700 (-0.10%)	-$9,427 (-1.32%)
Rescaled: Land & Sea scenario							
Very optimistic	9	93	667	-$496 (-0.07%)	-$1,081 (-0.15%)	-$126 (-0.02%)	-$1,699 (-0.24%)
Optimistic	10	105	750	-$558 (-0.08%)	-$1,216 (-0.17%)	-$142 (-0.02%)	-$1,911 (-0.27%)
Moderate	14	142	1,017	-$756 (-0.11%)	-$1,648 (-0.23%)	-$192 (-0.03%)	-$2,590 (-0.36%)
Pessimistic	29	289	2,067	-$1,537 (-0.22%)	-$3,351 (-0.47%)	-$391 (-0.05%)	-$5,266 (-0.74%)

[1] Based on Case Fatality Rates as in McKibbin *et al.* (11): Low=**0.023%**; Moderate=**0.23%**; Severe=**1.1667%**
[2] Sources: *Type I*: McKibbin *et al.* (11) and *Type II*: World Bank (12). Shocks are expressed in the original sources as a one-time, percentage of 2006 GDP shock. GDP used: Rupees 28.5 trillion (2006); US$/Rupees: 40 (2007)
[3] Labor shock based on direct costs: *moderate* mortality; and indirect costs: absenteeism of the ill and care-givers. Calculated in the Asia-Pacific G-Cubed global macroeconomic model as described in McKibbin *et al.* (11)
[4] Composite labor shock based on *severe* mortality rates (11), equal to -0.6%; and the shock due to illness and absenteeism of -0.8%. The value is for the region 'South Asia'. Given that India's GDP dominates this region, the use of the region (country-weighted) value as the country value can be excused for this Table.
[5] Shock based on decline in demand for a composite basket of goods, with a high weighting for services.
[6] A demand shock traced to efforts of individuals to avoid infection: reflecting reduced travel, consumption of services (hospitality, dining, theatres, etc). All Type II shocks are calculated in the 'Linkage' global CGE model.

The scaling in Table 13 is intended to yield indicative range of values for health and economic shocks. The human suffering due to severe sickness not leading to death is not

completely quantifiable, and imputing a monetary value to the deaths is not attempted. Critiques of the underlying macroeconomic and CGE models that yield the Type-I and II estimates in Table 13 can be read in the cited studies (11, 12).

Given the caveat, Table 13 indicates that the weighted scenario-based approach in this study yields much lower total health and economic shocks than in fixed, non-age-structured attack rate approaches. The shaded rows in Table 13 for 'moderate' epidemic size under the Air or Land & Sea entry situations are used for discussion hereafter. Under moderate case fatality rates (shaded column, Table 13), India could expect 250,000 deaths if a global outbreak occurs near an international air travel destination with links to Indian cities (Air scenario); and 142,000 deaths if the outbreak occurs in a more remote area, allowing authorities time to shut down risky air travel such that the disease only enters later through Land & Sea points. These can be compared to the projected 737,000-860,000 deaths under the two prior fixed attack rate analyses (11, 12). These mortality estimates are based on the case fatality rates (CFR) used in the McKibbin *et al.* study (11), which derived them from past pandemics. The moderate case fatality rate – captioned here as 'moderate virulence' – is suggested by the 1957-58 pandemic, which is estimated to have had a CFR in the range 0.04%-0.27% (41, 42). The results on estimated fatalities in this study are hence only indicative as they are not adjusted for the country distribution of host-level complicating factors (35) that might yield higher or lower case fatality rates compared to the international average. In the 1918 pandemic, pandemic excess mortality rates as percentages of the population in Indian provinces were two-three times higher than in the Western world (20). Such host-level factors will be considered in a forthcoming study.

The large range of the estimates of lost economic output in Table 13 are the result of different economic models used in the two source studies as well as the different assumptions on the underlying attack rates and virulence (see notes 3-6, Table 13). For Type-I *labor shocks* which are based on moderate mortality, the lost output under the Air scenario (moderate epidemic size) is $1.32 billion and for the Land & Sea scenario (moderate epidemic size) it is $750 million, based on the 2006-07 Indian GDP at nominal exchange rates. Similarly for Type-I *demand shocks* and for the same scenarios, the lost output is $340 million, and $192 million respectively. Though the scaled economic shock values for the Worst Case scenario from Table 3 are not specifically reported here, they fall in the vicinity of the 'pessimistic epidemic size' using the Air scenario.

The results immediately above suggest that at *moderate virulence* and *moderate epidemic size*, the total value of lost output based on Type-I shocks ranges between $950 million (Land & Sea scenario) to $1.66 billion (Air scenario). A lower range across the two scenarios is applicable if the forecast is more optimistic, and it is assumed at the state health systems will be able to reduce attack rates using nonpharmaceutical measures. With *optimistic epidemic sizes* and *moderate virulence*, the total for Type-I shocks is between $700 million (Land & Sea) to $1.1 billion (Air). At optimistic levels, a future pandemic outbreak near an international transport hub with connections to India costs $400 million more in lost output and 100,000 additional fatalities, based on the indicative calculations above. The overall range of benefits of preventing air transmissions as suggested across these two epidemic size/virulence combinations is $400-710 million.

It is uncertain which of the future scenarios of disease transmission to India will be realized if a pandemic-capable strain does emerge and causes an outbreak sufficient for the WHO to declare a phase five pandemic alert. Given that transmission to India is near-certain if a global pandemic gets underway (a 'high probability, high impact' event), the three scenarios – Air, Land & Sea, and Worst Case – capture the possibilities in terms of how that transmission occurs, and the various epidemic sizes using weights and assumptions of public health response are realizations of uncertain dynamics of internal spread.

In the world as the present where a global pandemic is not underway, a pandemic influenza epidemic in India is a '*low probability, high impact*' event. Given this characterization, what is the optimal annual resource allocation for preparedness under a limited public health resource envelope as in India? The answer is at best indicative, and depends on both the probability and impact sides of the characterization. The 'impact' has been defined in the analysis above. However, there is the fundamental uncertainty if a pandemic-capable viral strain will emerge in 2008 based on the currently circulating avian influenza A H5N1. This uncertainty drives the probability part. If quantified, this probability would yield the *expected value* of the averted total economic costs in 2008, which can then be related to benefits of preparedness interventions.

Recently, 'prediction markets' (43) have been used to determine the odds for the occurrence of uncertain short-term events for which modeling cannot yield a probability estimate. A prediction market for various events pertaining to avian and pandemic influenza was recently set up at the University of Iowa (44) and provides odds for the event

that the WHO declares a phase four pandemic alert, i.e., recognizing small localized clusters of human-to-human transmission (45). The phase four alert is several steps away from a pandemic; hence the 'odds' of the pandemic event internationally, which poses a risk to India, are even lower. However, these phase four odds – imputed from the price and based on the knowledge of the experts involved – are the best available for the 'low probability' event. The option 'yes' for the proposition 'will the WHO will declare pandemic alert phase four by July 1, 2008' had a bid of 7 cents vs. 93 cents for the option 'no'. This implies – reading the prices as odds – a probability of 0.07 for a phase four pandemic alert recorded in the first half of 2008.

Using this probability as the risk of a pandemic in the calendar year 2008 (i.e., realization of the WHO pandemic alert phase six); for any Government of India intervention that is 100% effective in preventing the importation of infectives via air into the country, the *expected economic benefit* has the following range (EC = economic costs):

Minimum: $0.07 \times (EC_{Air} - EC_{Land\&Sea})$ Optimistic Epidemic size

Maximum: $0.07 \times (EC_{Air} - EC_{Land\&Sea})$ Moderate Epidemic size

Using the values from Table 13, the indicated resource allocation equals $28-50 million for proactive phase 5-trigger interventions launched in the year 2008.

The indicative resource allocation above could be compared to potential costs of the intervention, which include the direct costs of planning, administering, and enforcing the intervention for the government, the various direct and indirect costs to private and government entities in the international air transport sector, and the indirect costs to individual travelers in-country and abroad due to the disruptions in travel plans as well as to firms from the unexpected cancellation of business trips. Besides the problems in estimating these direct and indirect costs of the intervention (which has not been planned or defined at this point), the counterfactual is also difficult to construct: how many flights would operate in the absence of intervention, how many travelers would self-protect by limiting travel, etc. Additionally, in the absence of an intervention targeting transmission via air, and when global outbreaks occur near an international transport hub, it is plausible that air travel would be disrupted globally once the outbreaks spread and the prevalence of disease increases. Therefore, the expected benefits should be compared only to the additional costs imposed because the Indian government acts proactively, e.g., suspending risky traffic in the interregnum between the declarations of WHO pandemic phases five and

six (above the 'natural' disruptions after phase six). Further research into such costs specific to the Indian air travel sector are required to evaluate such policy interventions for highly infectious human communicable diseases such as SARS, influenza, etc.

There are no prior examples of costly *proactive* action to suspend air travel. There are examples of reactive and delayed actions. In the SARS crisis of 2003, the Indian national carrier Air India curtailed flights to Singapore and Hong Kong, for up to 12 days between April 9th to 21st, but began this intervention after cases per day had already peaked in Singapore, seven days after the WHO travel advisory for Hong Kong and Guangdong (dated April 2, 2003), and more than two weeks after the WHO recommended screening of air passengers for SARS (dated March 27th, 2003). Luckily, India avoided hosting a SARS infection on its soil. Given that symptomatic transmission screening of pandemic influenza-like symptoms is difficult (more false positives) and asymptomatic transmission is a possibility, an intervention such as suspension of air travel, with concomitant economic and personal costs, is more relevant.

The focus on proactive interventions and the assumption of 100% prevention of transmission of disease to India via air is the difference between this study and other recent evaluations of the role of international air travel restrictions on pandemic influenza control and mitigation. These latter studies assume that some transmission to the country has *already occurred* at the point at which cities impose bans on further air travel (8, 9). The most optimistic result from these modeling studies – which use an SEIR model for the epidemic in each interconnected city within a global network of air travel – is that sequential travel bans may delay the peak of the epidemic (29). This has value, as it allows more time for public health authorities to prepare administrative and pharmaceutical capacity for prophylaxis of susceptibles and treatment of cases, and as a result has the potential to lower the attack rate and alleviate suffering related to morbidity and mortality.

However, in this study it has been demonstrated that if transmission via air could be prevented with proactive, costly interventions, then this is the same as averting the Air scenario (unavoidable in the absence of intervention) with its higher internal risks of spread and higher epidemic sizes. In this with-intervention future, only the Land & Sea scenario and its lower epidemic size realizations are likely. This likelihood depends on the probability of the pandemic, which was 'guesstimated' using prediction markets. The expected economic benefits of the intervention should be recalculated as the odds in the

prediction markets shift. The more likely a pandemic becomes, the higher is the cost-effectiveness for proactive interventions in a given year.

Implications for Research into Pandemic Influenza Containment and Mitigation: The goal of outbreak containment is to prevent the development of widespread epidemics. If epidemics cannot be prevented or have already been realized, then the public health interest in mitigation is to limit the overall attack rate, reduce the rate of serious illnesses and fatalities at an attack rate, and to lower and delay the peak daily rate of cases so that local health systems are not overwhelmed and services for other diseases do not suffer.

Consider a future world where the pandemic alert phase five has already been declared. In the scenario-based analysis above, the prevention of the Air scenario in favor of a Land & Sea entry scenario means that the attack rates are lowered as well as a time interval gained in which public health authorities in India can prepare.

Similarly, across Figs. 8 and 9, the epidemic curve is compressed and shifted, achieving the same benefits for New Delhi as mentioned above for the country under the scenario-based analysis. However, the EpiFlex results also predict the timing of the epidemic peak and its variation in response to modeled parameters. The scenario-based analysis, being inherently non-epidemiological, cannot yield predictions related to an epidemic curve, even if it incorporates the variation across states in individual, geographic, and population level factors driving R_0.

The predicted epidemic sizes in Delhi using the scenario-based analyses (Table 11) are much lower than in Fig. 9. The EpiFlex scale model may predict a changed epidemic dynamic for pandemic influenza in an Indian metropolis that diverges from techniques using historical average-based attack rates calibrated to prior pandemics and inter-pandemic influenza. For exploring this proposition, the reproductive number (R) of the epidemic within the initially fully-susceptible population of 300,000 in New Delhi was estimated from the EpiFlex output for both Patterns I and II of contact infectivity. The estimation was done as a weighted average of the various recorded values for the *number of secondary infections caused by an infected host* (*Z*), where the weights were the numbers of hosts at a particular value of *Z*. These estimates are shown in Table 14 below.

Table 14: Estimates of R from EpiFlex Structured Metapopulation Model

Reference R_0 for scenario-based estimation	1.4	GAR: 13.4% - 29.7%*
Reproductive number R, Pattern I	1.79	GAR: 87.5%
Reproductive number R, Pattern II	1.77	GAR: 42%

* See Table 11   Also see Supplementary Materials for tabulation of *Z* and the no. of hosts.

From Table 14, as was expected, the estimated reproductive numbers in the EpiFlex structured metapopulation modeling are higher. Considering that the basic reproductive number is always slightly higher than R (see the related discussion in the *Survey of Methods* section) – though the precise increment is indeterminate – the R_0 suggested by these estimates is at least close to 1.8. The implication for the prior scenario-based analysis is that the reference value of $R_0=1.4$ may be too low for calculating countrywide epidemic size, at least for urban areas. Hence, the epidemic sizes that were reported may be too low, and may need to be doubled. For example, taking the values from Table 8 above (asymmetric urban-rural R_0), the epidemic size in India for the Air scenario, with $R_0=1.7$ in urban areas is 251 million, vs. 113 million with $R_0=1.4$.

Across the two models run in EpiFlex – corresponding to Figs. 8 and 9 – the R values are very close, yet the epidemic sizes are halved from Pattern I to II. A potential explanation lies in superspreading[13] phenomena, which vastly increase with increased contact infectivity, given a superspreading threshold of $Z \geq 7$. From Table 15 below, Pattern I (GAR: 87.5%) has five times as many superspreading events (SSE) as Pattern II (GAR: 42%). In lay terms, the possibility of certain infected individuals infecting very large numbers of other individuals rises as the possibility of infection per type of contact is raised. This occurs even if the weighted average of such numbers of secondary infection per host (*Z*) is similar across the two patterns of per contact infectivity. With a different perspective, the distribution of *Z* in the Pattern I output is more variable. This matches the observation elsewhere based on actual infectious disease epidemic data (e.g., SARS) that disease conditions with a high variability in *Z* are set up for explosive, high size epidemics (46).

Table 15: Superspreader Events (SSE) in EpiFlex Modeling for New Delhi (Z >= 7)

Demographic	**Pattern I**: No. of SSE (%)	**Pattern II**: No. of SSE (%)
Male Lower Middle Class workers	991 (69%)	220 (78.3%)
Female Lower Middle Class workers	369 (25.7%)	59 (21%)
Other demographic groups	76 (5.3%)	2 (0.7%)
Total events	1,436 (100%)	281 (100%)

---

[13] Generally, a superspreading event (SSE) is one in which an infected individual subsequently infects an 'unusually' large number of secondary cases. The criterion for setting the SSE threshold varies by disease and the observed distribution of *Z* (defined in the text) in an epidemic. Lloyd-Smith *et al*. (46) suggest that the threshold value of *Z* should be at least 'many more than the average number'.

The insights from comparing the Pattern I output from EpiFlex modeling to Pattern II suggest that conventional or classical SEIR models cannot predict situations where explosive growth in epidemics might occur due to superspreading phenomena. In summary, the higher range of estimates suggest that the EpiFlex results are not easily comparable to the analysis using the average-based attack rates at different R_0 values in the rest of this study. At the very least, they suggest that the choice of $R_0=1.4$ as the reference level for this study was an optimistic one, at least for urban areas.

Further research is required to confirm the insights into the importance of superspreading, and that of per contact infectivity in driving superspreading, as noted above. It may be necessary to experiment with both SEIR and spatial-stochastic modeling techniques for pandemic influenza epidemics in Indian cities, state(s), or a scaled country-wide population. Results from these alternative methodologies can determine whether the results are plausible and comparable from both scenario-based and the metapopulation models. For spatial-stochastic modeling using 'explicit spatial' techniques, high-quality geographically linked demographic and land use data is now being collected for many states. Such data for New Delhi is currently available from public and private sources, and analyses as in recent studies (13-15) can be performed, though at higher resource and computational cost compared to the techniques in this study. This would allow the testing of spatially-defined containment interventions such as ring prophylaxis, quarantines, and social distancing, in the specific demographic and geographic context of India.

For improvements to the structured metapopulation model of New Delhi, it would be helpful to estimate contact rate patterns from direct observation at the various locations modeled (see Supplementary Materials). Household sizes here were only roughly estimated, as were the approximate sizes of individual mixing groups at various locations. At higher population scales, it would be needed to get better estimates of the sizes of hourly mixing density and rates at the locations from direct observation (with seasonal variations recorded if possible). Little has been formally recorded about influenza viral survivability given seasonal temperature and humidity patterns in India. These issues would impact transmission from mucosa, fomites, and aerosols (either on surfaces or directly from person-to-person, respectively), and also cause variation in epidemic severity when the beginning point is in different seasons.

For specific policy insights arising from the higher attack rates in the metapopulation model as implemented in Figs. 8-9 above, there is cause for concern since a very limited infection seed of five infections every five hours for fifteen hours, yields a very rapid and large epidemic (above 40% attack rate) in a metapopulation of New Delhi, even with moderate infectivity rates per contact. The epidemic size halves if low contact infectivity is modeled with the same seed. Most of the modeled mixing locations used for the EpiFlex metapopulation of 300,000 New Delhi residents were designed to feature airborne or surface-mucosal contacts. The extreme sensitivity to the contact infectivity indicates that within the modeled locations, given high enough the contact rates, the propagation of disease is limited by contact infectivity rather than the density and frequency of the contacts in the network. This implies that the epidemic ceiling is set by infectiousness rather than contact rates – unlike in models of influenza transmission in the West – and this was, in hindsight, to be expected in dense urban environments such as New Delhi.

This has the strong implication that public health interventions which attempt to interrupt influenza transmission through creating artificial zones of low or zero contact rates may fail if there are superspreaders (46) or if there is considerable variability in contact rates in the population (highly skewed contact proportions across similar sets of locations). These phenomena are difficult to predict from average-based scenario analysis with calibrated attack rates as in this study. Potential nonpharmaceutical interventions aimed at reducing contact rates as in Table 2 may be cost-effective – where cost is measured in terms of lost output and social inconvenience – only if they are universally adopted across location types and efficient. Public health communications that are persuasive and backed by believable governance, such as encouraging hand sanitizing, targeting behavioral modifications (related to sneezing, coughing, and personal hygiene), and decontaminating public spaces, may be more useful as they directly impact contact infectivity by type of contact.

Further research using EpiFlex is required to yield dominant policy insights across various scale population sizes, demographic profiles, and vector seeds, in terms of identifying the locations that are the richest breeding grounds for infections. Given multiple interconnected metapopulation models suited to metropolis, small town, and rural areas, such research could yield valuable insights for public health control measures in the event of a global pandemic influenza alert applicable to India. This research would also be valuable for other developing country settings. Additionally, by experimenting with

different contact skew patterns with a particular scheme of mixing locations, it can be determined if superspreader phenomena exist in the metapopulation modeling structures appropriate to mimic the reality of infectious disease dynamics in India. Any modeled evidence for the importance of such individual-level variation as in Z in densely populated urban and peri-urban areas would predict a break from the insights of classical SEIR models, and additionally validate the higher attack rates in EpiFlex. Evidence supporting the insights from the New Delhi scale model in this study would also suggest that Indian epidemics of pandemic influenza suffer higher attack rates than expected in the West, correlating with the back-projected attack rates from the unusually high deaths in Indian provinces during the pandemic of 1918-19.

Implications of this Research for Other Infectious Diseases: Pandemic influenza presents a specific case of highly infectious disease with high virulence that can rapidly spread across a country and across diverse transmission settings. Therefore, the findings in this study are generalizable only to a class of debilitating human communicable diseases with the capacity to cause widespread epidemics, i.e., retaining infectiousness without a trade-off in terms of lower virulence. In contrast, emerging infectious diseases of concern such as the hemorrhagic fevers Ebola and Marburg, have a lower R_0 since they depend on close contacts of infected and susceptibles (since the virus is transmissible only in body fluids and secretions) but are highly virulent. Given this, the findings in this study are applicable to respiratory diseases such as SARS, and smallpox. However, unlike these diseases, the possibility of asymptomatic transmission (21) for pandemic influenza renders it a particularly challenging future case for public health practice in developing countries without access to large stockpiles of pharmaceuticals.

References

1. World Health Organization. Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO. 2007 [cited 12/28/2007]; Available from: www.who.int/csr/disease/avian_influenza/country/cases_table_2007_10_12/en/index.html
2. Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solorzano A, Pappas C, et al. A Two-Amino Acid Change in the Hemagglutinin of the 1918 Influenza Virus Abolishes Transmission. Science. 2007 February 2, 2007;315(5812):655-9.
3. Centers for Disease Control. CDC Reproductive Health: Epi Glossary. CDC 2007 [cited 10/12/2007]; Available from: www.cdc.gov/reproductivehealth/EpiGlossary/glossary.htm
4. Bowen JT, Laroe C. Airline networks and the international diffusion of severe acute respiratory syndrome (SARS). Geographical Journal. 2006;172(2):130-44.
5. Grais RF, Ellis JH, Glass GE. Assessing the impact of airline travel on the geographic spread of pandemic influenza. European Journal of Epidemiology. 2003;18:1065-72.
6. Brownstein JS, Wolfe CJ, Mandl KD. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. PLoS Med. 2006;3(10):e401.
7. Hollingsworth TD, Ferguson NM, Anderson RM. Will travel restrictions control the international spread of pandemic influenza. Nat Med. 2006;12(5):497-9.
8. Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. Delaying the international spread of pandemic influenza. PLoS Medicine. 2006;3(6):0845-55.
9. Epstein JM. Controlling pandemic flu: the value of international air travel restrictions. PLoS ONE. 2007;2(5):e401.
10. Fox M. Bird flu mutating into human-unfriendly form. Reuters 2007 [cited 10/12/2007]; Available from: www.reuters.com/article/healthNews/idUSN0439354420071005
11. McKibbin W, Sidorenko AA. Global Macroeconomic Consequences of Pandemic Influenza. Sydney: Lowy Institute for International Policy; 2006.
12. World Bank. Chapter 1: Prospects for the Global Economy. Global Development Finance - The Development Potential of Surging Capital Flows. Washington, DC: The World Bank; 2006.
13. Ferguson N, Cummings D, Cauchemez S, Fraser C, Riley S. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature. 2005;437:209-14.
14. Longini IM, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DAT, et al. Containing pandemic influenza at the source. Science. 2005;309:1083-7.
15. Germann TC, Kadau K, Longini Jr IM, Macken CA. Mitigation strategies for pandemic influenza in the United States. PNAS. 2006;103(15):5935-40.
16. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006;442(7101):448-52.
17. Lilienfeld AM, Lilienfeld DE. Foundations of Epidemiology. Second ed. New York: Oxford University Press; 1980.
18. Chowell G, Ammon CE, Hengartner NW, Hyman JM. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. Journal of Theoretical Biology. 2006;241(2):193-204.
19. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review. 2000;42(4):599-653.
20. Murray CJ, Lopez AD, Chin B, Feehan D, Hill KH. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-20 pandemic: a quantitative analysis. The Lancet. 2006;368(December 23/30).
21. Bell DM. World Health Organization Writing Group. Nonpharmaceutical interventions for pandemic influenza, international measures. Emerg Infect Dis. 2006;12(1):81-7.
22. Watts DJ, Muhamad R, Medina DC, Dodds PS. Multiscale, resurgent epidemics in a hierarchical metapopulation model. PNAS. 2005 August 9, 2005;102(32):11157-62.
23. Stroud P, Valle SD, Sydroriak S, Riese J, Mniszweski S. Spatial dynamics of pandemic influenza in a massive artificial society. Journal of Artificial Societies and Social Simulation. 2007;10(4):9.
24. Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004;429(6988):180-4.

25. Hanley B. An object simulation model for modeling hypothetical disease epidemics–EpiFlex. Theoretical Biology and Medical Modelling. 2006;3(32).
26. Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. PNAS. 2004;101(42):15124-9.
27. Klein E, Laximinarayan R, Smith DL, Gilligan CA. Economic incentives and mathematical models of disease. Environment and Development Economics. 2007;12:707:32.
28. Institute of Medicine. Modeling Community Containment for Pandemic Influenza: A Letter Report: The National Academies Press; 2006.
29. Dutta A. The effectiveness of policies to control a human influenza pandemic: a literature review. Policy Working Paper. Washington, D.C.: The World Bank; 2008.
30. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. PNAS. 2004;101(16):6146-51.
31. Wu JT, Riley S, Fraser C, Leung GM. Reducing the impact of the next influenza pandemic using household-based public health interventions. Plos Medicine. 2006;3(9):e361.
32. Hatchett RJ, Mecher CE, Lipsitch M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. PNAS. 2007 May 1, 2007;104(18):7582-7.
33. Bootsma MCJ, Ferguson NM. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. PNAS. 2007;104(18):7588-93.
34. Bloom E, Wit Vd, Jose MJC-S. Potential economic impact of an avian flu pandemic on Asia. Manila: Asian Development Bank; 2005.
35. Meltzer MI, Cox NJ, Fukuda K. The economic impact of pandemic influenza in the United States: priorities for intervention. Emerg Infect Dis. 1999;5(5):659-71.
36. Doyle A, Bonmarin I, Lévy-Bruhl D, Strat YL, Desenclos JC. Influenza pandemic preparedness in France: modelling the impact of interventions. Journal of Epidemiology & Community Health. 2006;60(5):399-404.
37. Genugten MLL, Heijnen MLA, Jager JC. Scenario-analysis of expected number of hospitalisations and deaths due to pandemic influenza in the Netherlands. 2002.
38. Sertsou G, Wilson N, Baker M, Nelson P, Roberts MG. Key transmission parameters of an institutional outbreak during the 1918 influenza pandemic estimated by mathematical modelling. Theoretical Biology and Medical Modelling. 2006;3(1):38.
39. Rvachev L, Longini I. A mathematical model for the global spread of influenza. Mathematical Biosciences. 1985;75:3:22.
40. Kilbourne ED. Influenza pandemics of the 20th century. Emerg Infect Dis. 2006;12(1):9-14.
41. Beveridge WIB. The chronicle of influenza epidemics. Historical and Philosophic Life Sciences. 1991;13:223-35.
42. Simonsen L, Clarke MJ, Schonberger LB, Arden NH, Cox NJ, Fukuda K. Pandemic versus epidemic influenza mortality: a pattern of changing age distribution. The Journal of Infectious Diseases. 1998;178:53-60.
43. Wolfers J, Zitzewitz E. Prediction Markets. The Journal of Economic Perspectives. 2004;18(2):107-26.
44. University of Iowa Health Prediction Markets. Avian Influenza Prediction Market. 2007 [cited 12/12/2007]; Available from: http://fluprediction.uiowa.edu/fluhome/Market_AvianInfluenza.html
45. Centers for Disease Control. Pandemic Phases. [cited 12/12/2007]; Available from: http://www.cdc.gov/flu/pandemic/phases.htm
46. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the impact of individual variation on disease emergence. Nature. 2005;438:355–9.

Paper II: Potential Infections during a Pandemic of Influenza in the States of India: Implications for Public Health Policy and Research
Supporting Material (Appendix)

1. Demographics of India

Urban Population in 2006, '000

States	0-4	5-9	10-14	15-19	20-44	45-64	65+
Andaman & Nicobar	12	13	14	15	60	26	9
Andhra Pradesh	1,932	2,065	2,287	2,332	8,815	3,597	1,199
Arunachal Pradesh	25	33	35	35	121	42	13
Assam	422	454	466	442	1,522	534	147
Bihar	1,152	1,266	1,238	1,076	3,132	1,266	390
Chandigarh	94	97	104	107	388	142	58
Chhattisgarh	550	560	550	521	1,748	706	224
Delhi	1,240	1,452	1,513	1,618	6,625	2,103	575
Goa	75	78	80	83	319	123	49
Gujarat	2,092	2,114	2,178	2,178	8,455	3,288	1,046
Haryana	743	772	808	815	2,818	947	379
Himachal Pradesh	58	62	65	69	266	108	44
Jammu & Kashmir	291	279	316	330	1,076	401	130
Jharkhand	727	820	820	760	2,347	934	240
Karnataka	1,761	1,881	2,001	2,081	7,965	3,242	1,101
Kerala	668	694	694	737	3,460	1,627	677
Madhya Pradesh	2,174	2,138	2,102	1,977	6,433	2,336	809
Maharashtra	4,320	4,506	4,599	4,785	18,350	7,061	2,834
Manipur	51	65	69	71	241	83	25
Meghalaya	42	54	57	58	198	68	21
Mizoram	41	52	55	57	194	67	20
Nagaland	31	39	42	43	146	50	15
Orissa	573	617	654	641	2,374	962	351
Pondicherry	60	59	64	68	305	140	49
Punjab	831	840	934	991	3,747	1,472	623
Rajasthan	1,738	1,826	1,811	1,649	5,124	1,900	663
Sikkim	6	8	8	8	28	10	3
Tamil Nadu	2,565	2,533	2,757	2,918	13,146	6,028	2,116
Tripura	51	66	70	72	245	84	26
Uttar Pradesh	5,089	4,854	4,893	4,502	13,075	5,011	1,762
Uttaranchal	276	264	279	281	901	356	134
West Bengal	2,068	2,381	2,550	2,525	9,525	3,776	1,179
INDIA	33,426	34,390	35,354	34,390	118,918	47,567	15,106

Source: Census of India 2001, projections

Rural Population in 2006, '000

States	0-4	5-9	10-14	15-19	20-44	45-64	65+
Andaman & Nicobar	22	24	26	26	108	46	15
Andhra Pradesh	5,090	5,441	6,026	6,143	23,227	9,478	3,159
Arunachal Pradesh	73	94	99	101	346	119	36
Assam	2,616	2,814	2,888	2,740	9,429	3,307	913
Bihar	9,829	10,804	10,560	9,179	26,725	10,804	3,331
Chandigarh	11	11	12	12	44	16	7
Chhattisgarh	2,003	2,038	2,003	1,897	6,363	2,570	815
Delhi	73	86	90	96	392	124	34
Goa	64	66	68	70	270	104	42
Gujarat	3,296	3,329	3,430	3,430	13,317	5,179	1,648
Haryana	1,635	1,699	1,780	1,796	6,205	2,084	834
Himachal Pradesh	503	532	561	596	2,290	925	382
Jammu & Kashmir	836	804	909	950	3,093	1,153	373
Jharkhand	2,467	2,783	2,783	2,580	7,966	3,168	815
Karnataka	3,190	3,407	3,625	3,770	14,426	5,872	1,994
Kerala	1,927	2,001	2,001	2,124	9,979	4,693	1,951
Madhya Pradesh	5,859	5,762	5,665	5,326	17,335	6,295	2,179
Maharashtra	5,426	5,660	5,776	6,010	23,047	8,869	3,559
Manipur	143	184	194	199	680	234	72
Meghalaya	166	213	225	231	788	270	83
Mizoram	39	50	52	54	184	63	19
Nagaland	147	189	200	205	700	240	74
Orissa	3,043	3,272	3,468	3,403	12,598	5,104	1,865
Pondicherry	28	28	30	32	145	66	23
Punjab	1,463	1,479	1,645	1,745	6,598	2,593	1,097
Rajasthan	5,611	5,896	5,849	5,326	16,548	6,134	2,140
Sikkim	43	55	58	59	202	69	21
Tamil Nadu	2,646	2,613	2,844	3,010	13,560	6,218	2,183
Tripura	235	302	319	327	1,115	383	117
Uttar Pradesh	18,738	17,873	18,017	16,576	48,141	18,449	6,486
Uttaranchal	747	713	754	761	2,437	963	363
West Bengal	5,260	6,055	6,483	6,422	24,221	9,603	2,997
INDIA	82,242	84,614	86,986	84,614	292,591	117,036	37,167

Source: Census of India 2001, projections

2. Demographic/census age groups and attack rate age groups

Age groups in Longini *et al.*	Age groups in census data	Match for applying attack rates
<1	0-4	Average of attack rates for <1 and 1-5 used.
1-5		
6-10	5-9	Considered same
11-14	10-14	Considered same
15-17	15-19	Considered same
18-44	20-44	Considered same
45-64	45-64	Exact
65+	65+	Exact

3. Scenario-based analysis: Geographic and health system weights

Weights based on the index of geographic risk

States	Air Scenario		Land & Sea Scenario		Worst Case Scenario	
	Urban	Rural	Urban	Rural	Urban	Rural
Andaman & Nicobar	Low	Lowest	Medium	Lowest	Medium	Lowest
Andhra Pradesh	Medium	Medium	Low	Lowest	Medium	Medium
Arunachal Pradesh	Lowest	Lowest	Low	Lowest	Low	Lowest
Assam	Low	Low	Low	Low	Low	Low
Bihar	Low	Low	Lowest	Lowest	Low	Low
Chandigarh	Low	Low	Lowest	Lowest	Low	Low
Chhattisgarh	Lowest	Lowest	Lowest	Lowest	Lowest	Lowest
Delhi	High	High	Low	Lowest	High	High
Goa	Low	Low	Low	Low	Low	Low
Gujarat	Medium	Low	Medium	Lowest	Medium	Low
Haryana	Medium	Medium	Lowest	Lowest	Medium	Medium
Himachal Pradesh	Lowest	Lowest	Lowest	Lowest	Lowest	Lowest
Jammu & Kashmir	Lowest	Lowest	Lowest	Lowest	Lowest	Low
Jharkhand	Lowest	Lowest	Lowest	Lowest	Lowest	Lowest
Karnataka	High	Medium	Low	Lowest	High	Medium
Kerala	Medium	Medium	Low	Low	Medium	Medium
Madhya Pradesh	Low	Lowest	Lowest	Lowest	Low	Lowest
Maharashtra	High	High	High	Low	High	High
Manipur	Low	Lowest	High	High	High	High
Meghalaya	Lowest	Lowest	Medium	Medium	Medium	Medium
Mizoram	Lowest	Lowest	High	High	High	High
Nagaland	Lowest	Lowest	High	High	High	High
Orissa	Low	Lowest	Low	Lowest	Low	Low
Pondicherry	Medium	High	Low	Lowest	Medium	High
Punjab	Low	Lowest	Lowest	Lowest	Low	Lowest
Rajasthan	Low	Low	Lowest	Lowest	Low	Low
Sikkim	Lowest	Lowest	Lowest	Lowest	Lowest	Lowest
Tamil Nadu	High	High	Medium	Low	High	High
Tripura	Low	Lowest	High	High	High	High
Uttar Pradesh	Low	Medium	Lowest	Lowest	Low	Medium
Uttaranchal	Lowest	Lowest	Lowest	Lowest	Lowest	Lowest
West Bengal	High	High	Medium	Lowest	High	High

Index of geographical risk weights, Land & Sea scenario, for urban populations of states

3.3 Index of geographical risk weights, Worst Case scenario, for urban populations of states

3.4 Weights based on the index of state health system preparedness

States	Weight	States	Weight
Andaman & Nicobar	Lowest	Madhya Pradesh	Medium
Andhra Pradesh	Medium	Maharashtra	High
Arunachal Pradesh	Lowest	Manipur	Low
Assam	Low	Meghalaya	Lowest
Bihar	Low	Mizoram	Medium
Chandigarh	Lowest	Nagaland	Low
Chhattisgarh	Low	Orissa	Low
Delhi	Lowest	Pondicherry	Lowest
Goa	Low	Punjab	Medium
Gujarat	Medium	Rajasthan	Medium
Haryana	Lowest	Sikkim	Lowest
Himachal Pradesh	Medium	Tamil Nadu	High
Jammu & Kashmir	Low	Tripura	Lowest
Jharkhand	Low	Uttar Pradesh	Medium
Karnataka	Medium	Uttaranchal	Lowest
Kerala	High	West Bengal	High

4. Background data by state: Index of geographical risk

Urban populations

States	Urban Persons/km²	Av. monthly intl. passenger movements*	Av. monthly domestic passenger movements*	National Highway Kilometers**
Andaman & Nicobar	6,732	0	16,800	300
Andhra Pradesh	3,960	82,290	277,920	4,472
Arunachal Pradesh	13	0	29	392
Assam	3,566	810	77,910	2,836
Bihar	5,379	0	25,680	3,537
Chandigarh	8,198	0	10,800	24
Chhattisgarh	154	0	11,100	2,184
Delhi	18,175	473,910	860,400	72
Goa	1,437	33,270	104,220	269
Gujarat	3397	37,110	180,000	2,871
Haryana	5,392	0	127,628	1,468
Himachal Pradesh	1,955	0	1350	1,208
Jammu & Kashmir	556	0	72,960	823
Jharkhand	338	0	807	1,805
Karnataka	3,819	70,890	421,200.	3,843
Kerala	2,497	240,450	102,420	1,440
Madhya Pradesh	2,550	0	19,110	5,200
Maharashtra	6,008	554,730	110,7720	4,176
Manipur	4,359	0	11,070	959
Meghalaya	2,808	0	2,238	810
Mizoram	824	0	4,470	927
Nagaland	2,325	0	1,890	494
Orissa	1,939	0	18,180	3,704
Pondicherry	4,409	0	203,085	53
Punjab	4,993	33,630	6,420	1,557
Rajasthan	2,649	4,320	58,290	5,585
Sikkim	76	0	340	62
Tamil Nadu	3,455	225,630	406,170	4,183
Tripura	3,331	0	19,470	400
Uttar Pradesh	7,257	9,660	56,970	5,599
Uttaranchal	158	0	1,890	1,991
West Bengal	7,160	60,990	319,230	2,325

Data sources:
* "India Air Transport Statistics 2005-06", Directorate General of Civil Aviation, Government of India
** National Human Development Report 2001, Planning Commission, Government of India (latest available)
Population density numbers were last calculated for all states in the Census of India, 1991

Index of geographical risk: background data contd.: **Rural populations**

States	Rural Persons/km²	Road length per million pop. 1997**	Av. Monthly domestic passenger movements*	National Highway Kms.**	Av. % village connected, 1996-97**
Andaman & Nicobar	32	32.93	16,800	300	78.13
Andhra Pradesh	206	24.25	277,920	4,472	89.13
Arunachal Pradesh	13	128.11	29	392	70.51
Assam	305	27.04	77,910	2,836	88.87
Bihar	782	9.28	25,680	3,537	58.38
Chandigarh	2,042	21.91	10,800	24	100.00
Chhattisgarh	154	15.80	11,100	2,184	57.38
Delhi	1,750	21.27	860,400	72	100.00
Goa	240	57.09	104,220	269	99.80
Gujarat	174	19.59	180,000	2,871	95.58
Haryana	369	15.65	127,628	1,468	99.18
Himachal Pradesh	100	48.70	1350	1,208	57.59
Jammu & Kashmir	46	23.06	72,960	823	75.26
Jharkhand	338	9.28	807	1,805	53.17
Karnataka	195	28.63	421,200.	3,843	99.79
Kerala	660	46.26	102,420	1,440	94.23
Madhya Pradesh	154	26.33	19,110	5,200	57.75
Maharashtra	197	40.98	110,7720	4,176	82.66
Manipur	75	47.57	11,070	959	64.97
Meghalaya	85	38.55	2,238	810	65.10
Mizoram	23	53.66	4,470	927	93.61
Nagaland	100	122.37	1,890	494	95.03
Orissa	208	75.27	18,180	3,704	68.11
Pondicherry	1,000	24.05	203,085	53	100.00
Punjab	350	27.04	6,420	1,557	98.48
Rajasthan	129	25.43	58,290	5,585	69.65
Sikkim	76	36.68	340	62	89.04
Tamil Nadu	332	34.25	406,170	4,183	76.66
Tripura	262	43.32	19,470	400	82.71
Uttar Pradesh	569	15.90	56,970	5,599	47.27
Uttaranchal	158	9.54	1,890	1,991	46.45
West Bengal	678	9.91	319,230	2,325	58.16

Data sources:
* "India Air Transport Statistics 2005-06", Directorate General of Civil Aviation, Government of India
** National Human Development Report 2001, Planning Commission, Government of India (latest available)
Population density numbers for rural and urban areas of all states were last calculated for all states in the Census of India, 1991. Some states' population densities were calculated for 2001.

5. Background data by state: Index of state health system preparedness

State	Per capita public health expenditure** Rs. 2001-02	Number of District Hospitals* (2007)	Number of Sub Div. Hospitals* (2007)	No. of SCs* (2007)	Number of PHCs* (2007)	Number of CHCs* (2007)	IMR* (SRS 2005)
A&N Islands	1,228	2	0	108	20	4	27
Andhra Pradesh	182	20	56	12,522	1,570	167	57
Arunachal Pradesh	627	8	0	379	85	31	37
Assam	176	20	3	5,109	610	100	68
Bihar	92	26	22	8,858	1,641	70	61
Chandigarh	804	1	0	13	0	1	19
Chhattisgarh	121	16	0	4,692	518	118	63
Delhi	426	9	14	41	8	0	35
Goa	685	2	0	172	19	5	16
Gujarat	147	23	23	7,274	1072	273	54
Haryana	163	21	13	2,433	408	82	60
Himachal Pradesh	493	12	36	2,069	439	66	49
Jammu & Kashmir	271	14	26	1,888	374	80	50
Jharkhand	146	22	16	3,958	330	194	50
Karnataka	206	24	152	8,143	1,679	254	50
Kerala	240	14	41	5,094	909	107	14
Madhya Pradesh	132	47	55	8,874	1,192	229	76
Maharashtra	196	35	81	10,453	1,800	407	36
Manipur	345	7	0	4,20	72	16	13
Meghalaya	407	3	1	4,01	101	25	49
Mizoram	836	8	1	3,66	57	9	20
Nagaland	414	11	0	3,97	84	21	18
Orissa	134	32	22	5,927	1,279	231	75
Pondicherry	841	4	0	77	39	4	28
Punjab	258	20	55	2,858	484	126	44
Rajasthan	182	28	144	10,512	1,713	325	68
Sikkim	825	4	0	147	24	4	30
Tamil Nadu	202	29	270	8,683	1,252	165	37
Tripura	301	2	11	539	73	10	31
Uttar Pradesh	84	74	0	20,521	3,660	386	73
Uttaranchal	178	16	18	1,631	222	49	42
West Bengal	181	15	79	10,356	922	346	38

Abbreviations:
SC: Health sub-center; PHC: Primary healthcare center; CHC: Community healthcare center
SRS: Sample Rural Survey

Data sources:
* Status Report, June 2007, National Rural Health Mission, Government of India
** National Health Accounts 2002, Government of India (latest available for all states)

6. **Calculation of the indices**

Raw value of the index of geographical risk (urban) =
(Av. monthly international air passenger movements +
0.5× Average monthly domestic passengers movements +
0.5× National highway kilometers in state) × Urban population density
per square kilometer

Raw value of the index of geographical risk (rural) =
[International airports factor ×
(Road length per million population × Av. percentage villages connected by road +
0.25× National highway kilometers in state) +
0.25× Average monthly domestic passengers movements]× Rural population density
per square kilometer

Raw value of the index of state health system preparedness =
[Per capita health expenditures ×
 (2 × No. of district hospitals +
  1.5 × No. of sub-divisional hospitals +
  No. of health sub-centers +
  No. of primary healthcare centers +
  0.5 × No. of community healthcare centers)]/Infant mortality rate per 1000 births

Transformation of the raw values (all indices)

Transformed index value for state i = $\dfrac{(I_{ri} - \frac{1}{N}\sum\limits_{i=1}^{N} I_{ri})}{\sigma_{Ir}}$

where I_{ri} is the **raw** value of the index (as above) for state i, N is the number of states, and σ_{Ir} is the standard deviation of the series of N raw index values

Final index value = *If* Transformed index value > 1, then 1; *else*
Transformed index value + 1

7. **Weight label assignments**

Based on *percentile distributions* of the 'final index values' above for each index:

Weight label for state i	Index of geographical risk		Index of state health system preparedness
	Urban	Rural	
Lowest	< 30th percentile	< 50th percentile	< 30th percentile
Low	> 30th percentile	> 50th percentile	> 30th percentile
Medium	> 70th percentile	> 70th percentile	> 60th percentile
High	≥ 90th percentile	≥ 90th percentile	≥ 90th percentile

8. **EpiFlex Structured Metapopulation Model**

Aim: Rapidly generate *N* trials of an EpiFlex simulation of an epidemic of H5N1 human influenza in New Delhi (India), using a scale model of the population (specifics below). The target results – to be calculated from the model output of immune/removed – are the average total morbidity and mortality rates. The overall aim is to provide a quick spatial-stochastic modeling estimate to compare with estimates of total morbidity and mortality from applying age-structured gross attack rate patterns from prior published research to the demographics of New Delhi in 2006.

Data available for EpiFlex modeling: Age-structured demographics for New Delhi (five age bins from 0-4 to 65+); demographic splits, no gender, as % of working age population for agricultural workers, informal workers, and formal sector workers (service and manufacturing), airport passengers per annum and per day (domestic, foreign)

Guesstimates: % of housewives in adult females (i.e., housebound, no formal work); % of blue-collar vs. white collar in working age population, no gender, (i.e., split within formal sector workers); airport workers (number)

Hardware: Intel Core Duo T2250 Laptop (2x1.73GHz processors, 2MB L2 Cache, 533 MHz FSB)

Model Particulars:
A full description of the EpiFlex modeling methodology and definitions of the terms used here is given elsewhere (25). Movement cycles for all the different demographic groups are omitted here.

Model run: 450 cycles, 5 cycles per day

Population as a scale model: 300,000 New Delhi inhabitants, distributed in demographic bins as per 2001 census (estimated 2006 population: 15.6 million)

Disease definition: Pandemic Influenza A
Case fatality rate: 1.3%
Manifesting period infectivity: 100% of the probability of infection per contact (as defined below)
Incubating fractional infectivity: 1% of Manifesting   Incubating period: 1-4 days
Prodromal fractional infectivity: 20% of Manifesting   Prodromal period: 2-4 days
                                                             Manifesting period (ill): 2-3 days
Mobility, manifesting period: 3-Medium behavior/2-Major impairment (prostration)
Mobility, prodromal period: 2-Medium (serious cold)/1- Minor impairment (some fatigue)
As defined in Table 10 of the main text, contact infectivity patterns are:

	Probability of transmitting infection per contact (%)	
Contact type	Pattern I	Pattern II
Close airborne (aerosol)	4.5%	1.2%
Casual airborne (aerosol)	2.5%	1%
Surface-to-hand (mucosa)	1.5%	1.2%
Food-to-hand (mucosa)	1.2%	1.2%

EpiFlex Area: Delhi Urban (94% of population as per census)

EpiFlex Groups: Adult Male Lower Middle Class (LMC) Worker, Adult Female LMC Worker, Adult Male Middle Class/Upper Middle Class (MC/UMC) Worker, Adult Female MC/UMC Worker, Adult Male Slum Worker, Adult Female Slum Worker, College students, Students Class I-V, Students Class VI-VIII, Students Secondary and Sr. Secondary, Medical Staff, Nonworking females, Other nonworking dependents = *14 groups*

EpiFlex Locations (number of cells, contact rate skew): Home (X, high), airport (1, high), school (X, medium), local transport hub (X, high), college (X, high), workplace LMC (X, medium), workplace MC/UMC (X, medium), slum (X, medium), hospital (X, medium), clinic (X, medium), movie theatre (12, high), bar (30, medium)
X: model calculates the number of cells at runtime.
Location 'home' is considered a grouping of individual nuclear family dwellings, or a neighborhood, comprising 30-40 individuals.

EpiFlex Cycle: 5 movement cycles per day (e.g., home, local transport hub, workplace LMC, local transport hub, home). Each cycle roughly corresponds to about 4.5 hours

EpiFlex H5N1 Influenza Initiating Infection Vector: 3 cycles at the location 'home', forced infections per cycle = 5 (i.e. 15 infections are forced in the city)

Location contact rate definitions:

Location	Random Cell Exchange Function	Contact rate by type at this location *per cycle*	Skew Fn. Root*
Airport	0.0001	Air-casual 70, Surface-hand-mucosa: 15	0.482
Bar	0.004	Air-casual 15, Surface-hand-mucosa: 10	0.881
Clinic	0.003	Air-close 5, Surface-hand-mucosa: 2	1.405
College	0.0008	Air-casual 5, Surface-hand-mucosa: 5	0.619
Home	0.007	Air-casual 5, Surface-hand-mucosa: 5	0.619
Hospital	0.006	Air-casual 20, Surface-hand-mucosa: 15	0.881
Local Transport Hub	0.01	Air-casual 40, Surface-hand-mucosa: 10	0.706
Movie Theatre	0.0001	Air-casual 10, Surface-hand-mucosa: 5	0.706
School	0.0001	Air-casual 5, Surface-hand-mucosa: 3	1.056
Slum	0.003	Air-casual 5, Surface-hand-mucosa: 3	0.881
Workplace LMC	0.003	Air-casual 10, Surface-hand-mucosa: 5	0.794
Workplace MC/UMC	0.002	Air-casual 5, Surface-hand-mucosa: 3	0.794

* $y = Root^2/(Root^2 + x^2)$ where y: fraction of contact frequency, x: human's queue position

Public health response:
Alert trigger: Fatalities appearance   Probability of noticing: 0.95
Number of occurrences to trigger alert: 5   Length of alert: 60 days
Post alert detection method: Symptoms
Detected fraction by phase: Incubation-1%;   Prodromal-60%;   Manifesting-80%
Contact mitigation (fractional infectivity after detection for contact type):
        Blood (stick/needle): 100%   Blood (oral/mucosal): 100%
        Close airborne: 60%   Casual airborne: 40%
        Sex: 5%   Surface-hand-mucosa: 50%
        Food contact: 50%

9. Epidemic Reproductive Number Estimation from EpiFlex Model Output

Z = Number of secondary infections caused by a single infected host.

Contact infectivity pattern I model, GAR: 87.5%

Z	No. of infected hosts displaying Z (weights)
0	115,059
1	80,080
2	36,532
3	15,211
4	6,446
5	2,931
6	1,394
7	747
8	339
9	174
10	19
11	17
12	8
13	1
14	1

Reproductive number, R: weighted average of Z from table above = **1.79***
(Excludes Z=0)

Contact infectivity pattern II model, GAR: 42%

Z	No. of infected hosts displaying Z (weights)
0	54,149
1	38,412
2	18,280
3	7,905
4	3,416
5	1,266
6	581
7	187
8	62
9	22
10	9
11	1

Reproductive number, R: weighted average of Z from table above = **1.77***
(Excludes Z=0)

References

1. Hanley B. An object simulation model for modeling hypothetical disease epidemics–EpiFlex. Theoretical Biology and Medical Modelling. 2006;3(32).

Paper III: A Comprehensive Look at Antiretroviral Therapy in India: Demand, Utilization, Impact of Non-Availability, and the Cost-Effectiveness of Different Treatment Options Including Second-line Therapy

Background: The estimated adult HIV prevalence in India, after accounting for biases, is 0.4% in 2008. Of the related 2.41 million HIV+ adults, about 0.82 million are likely to utilize ART across free/subsidized programs and those paying out-of-pocket (OOP). The availability of second-line combination ART remains very limited.

Objective: Estimate the spending on ART across public, NGO and out-of-pocket sectors. Project the mortalities and morbidities in those HIV+ not receiving ART. Estimate the cost-effectiveness associated with different treatment strategies in the context of a doubling of access to ARVs in the public sector. Inform government policy, especially for addressing treatment failure within the expanded first-line ART program. Validate a customizable and tractable model to estimate outcomes in ART cohorts for use in the field.

Methods: Based on prevalence, demand and utilization for ART in OOP and free/subsidized sectors were estimated, as well as total spending. A HIV state-transition model exploiting differences in mortality and opportunistic infection risk across CD4 cell-count strata was used to simulate the disease outcomes in the cohort of 1.6 million HIV+ individuals in India not receiving any ART, but with some receiving co-trimoxazole prophylaxis against opportunistic infection. The model was also used to simulate the effectiveness of single-line and two-line treatments for a cohort of 122,947 treatment-naïve individuals. Incremental cost effectiveness of the following ART strategies: physician-determined switching of regimen to a second-line on treatment failure (TF), continuing first-line ART till end of follow-up regardless of treatment failure, and stopping ART at treatment failure, were compared to no treatment and to each other. Early treatment start vs. later treatment start based on CD4-testing were incorporated and compared for all treatment models.

Results: All ART strategies were found cost-effective compared to no treatment. Co-trimoxazole reduced opportunistic disease incidence and related costs. Estimates are in line with more complex, less tractable models. Premature deaths averted ranged from 86,300 (stopping ART at first-line TF) to 99,340 (early starting, two-line ART). The use of the realistic treatment/adherence horizon of a 'follow-up' period, five and seven years for single- and two-line strategies respectively, did not change the rank order in ICER terms from other studies. Continuing first-line ART till end-of-follow-up even with viral resistance was cost effective compared to ending ART at first-line TF (ICER: US$1080-1110 per year of life saved). Starting ART earlier (CD4 < 250 cells/mm3) was generally more cost-effective. However, compared to other studies, a two-line strategy was not found incrementally cost-effective against continuing first-line ART till end of follow-up, by WHO standards.

Conclusions: Public sector share of spending on ART is low. Expanding first-line ART coverage is cost-effective, and will avert many premature deaths and reduce severe morbidity. Increasing survival with second-line ART comes at higher cost. Co-trimoxazole prophylaxis in situations without ART is an interim necessity. Continuing, compared to stopping, first-line ART in the case of treatment failure is promising, as long viral resistance and transmission is monitored and curtailed.

Keywords: antiretroviral therapy, second line, treatment failure, HIV/AIDS, India

Introduction

The health policy setting for HIV/AIDS in India has moved from 'crisis management to sustained strategic response', in line with developments in the global profile of the disease. This movement has two originating factors. One, the overall size of the epidemic in India has been drastically reduced on re-estimation of prevalence rates, which means that the country is not host to the world's highest (or second highest) number of people living with HIV/AIDS (PLWHA). Second, stabilization in incidence over the recent years, demonstrated in national and local seroprevalence surveys, has meant that the strategic priorities have shifted to implementing an enhanced care and treatment platform for those infected, while maintaining vigilance and responding with prevention efforts in new hotspots of incidence and for high-risk groups. Generally, treatment with most classes of antiretrovirals has been proven to make available some years of healthy life to patients, with mostly tolerated side-effects and drug toxicities, while reducing rates of premature mortality and debilitating morbidity from opportunistic infections that accompany advanced HIV disease.

The focus of this study is on antiretroviral (ARV) policy in India, especially with regards to the decision of the Government of India to triple the coverage of its free antiretroviral therapy (ART) program by 2011 to nearly 300,000 individuals [1]. The Government of India expansion program will still only reach a fraction of those who would be eligible for ART using WHO guidelines. The rest of those eligible would continue to purchase their drugs from private suppliers, paying out-of-pocket, or go without the treatment option, risking premature death and higher intermediate morbidity. For both illuminating the general landscape of benefits to treatment scale-up as well as being important public health facts in themselves, it is important to know the current population sizes of these three categories of individuals: those receiving ART of some form from free public or subsidized NGO sources; those who are purchasing ARVs using out-of-pocket funds; and those receiving no ART at all. The basic building block for the estimation of the size of these three groups by Indian state (province) is the point prevalence estimate of HIV+ individuals in India for 2008. This figure has been controversial, and even studies of ART policy published as recently as late 2007 have continued to cite the now superseded figure of 5.7 million PLWHA [2]. Therefore, this particular study represents the first comprehensive look at ART policy in the era of a significant prevalence re-estimation.

Two sources of data on HIV prevalence by state are available – prevalence in non-randomly selected sentinel surveillance sites from the 2006 national round, and rates from a randomly-selected population sample under the country's Third National Family Health Survey which gathered data during 2005-6. Combining the two sources allows the elimination of bias obtaining in each source in isolation, and allows the fixing of a point estimate with more precision than the 20% error margins usually reported [3].

After the by-state point prevalence for 2008 is estimated in the next section, the overall size of the three sub-groups within the total PLWHA are calculated using ART utilization rates for government, NGO, and private sector (i.e., out-of-pocket) providers. Based on market survey of prices prevailing in the public vs. private sector, the first two of the three D's of resource needs in treatment and care – drugs, diagnostics, and doctors [4] – can be sized in current dollars. These estimates yield insights into whether the government's current levels of spending are supportable, too high as a proportion of the total, or too low; given that cost-effectiveness (or lack of) for the treatment interventions is known. This implies that cost-effectiveness of the planned Indian treatment scale-up program, and all the intervention variations involved, also needs to be estimated.

Before the estimation of cost-effectiveness, the total deaths and episodes of severe morbidity are simulated in the entire cohort of eligible adult PLWHA who do not receive ART, using the numbers estimated from national prevalence and the utilization rates. This is conducted using an average-parameter based, non-stochastic, disease state-transition model, with HIV progress in individuals without ART patterned on published studies on the natural history of HIV in India, as well as other clinical data on opportunistic infection risk and mortality risk from resource-poor settings. The model is subsequently validated when it is used to calculate cost-effectiveness, by yielding economic valuations which are in a range with the results of more complex simulations of comparable treatment interventions.

Such recent studies include those which exploit a complex stochastic, individual-based clinical state transition simulation: the Cost-Effectiveness of Preventing AIDS Complications International Model (CEPAC), and have yielded such results for various treatment options in resource-poor settings [5, 6] , including India [2]. Among detailed results, they also confirmed that 'first-line' ART (the first drug regimen offered to patients who have never received antiretrovirals, usually a combination suited to the patient's medical history and specific HIV condition) is always cost-effective, and 'second-line' ART

(offered to those for whom first-line therapy no longer yields benefits) may also be under certain conditions. In India, the issues of treatment failure is of major concern with first-line drugs which are routinely prescribed, since their prices have seen substantial decline with the high volume and increasing diversity of generics production. This study re-estimates incremental cost-effectiveness of treatment strategies given the reality of treatment failure, using the same cohort simulation model noted above. Disease progress in individuals with ART, and hence ART efficacy, is patterned on published studies, including clinical trials data previously used in associated simulations. Representing a novel element to such calculations for India, policy-relevant cost-effectiveness comparison is made in the context of treatment failure, between switching to second-line based on physician-determined treatment failure vs. continuing with the first-line therapy (which yields some benefits). Previously analyzed in similar simulations, resource-sparing variations within the purview of these strategies are modeled, such as starting treatment in any type of ART model 'later', i.e., limiting ART initiation to patients with more advanced disease as proxied by lower CD4 T-cell counts. In a policy setting of limited access to second-line ARVs (and even some limit to subsidies for first-line ART), there is enhanced relevance to this selection of treatment models for simulation. An economic comparison based on averted deaths and years of life gained would be beneficial for informing policymakers on their choices.

An important product of this study is the design, implementation and validation of a customizable model of antiretroviral use in a cohort of individuals in resource-poor settings where the recently popularized 'public health approach' to ART [7-9] is appropriate. Briefly, the aspects of this detailed approach relevant to this discussion are standardized treatment protocols disseminated to physicians who may be hard-pressed or under-trained to handle every clinical profile presented; treating a large number of patients with the optimal resources across the three Ds, rather than treating with specialized care only those able to pay or at most need (this recognizes at least partially the epidemiological benefits of a level of well-adhered and structured ARV treatment regimen in the absence of a HIV vaccine, i.e., a beneficial externality similar to immunization); and using strategies to reduce costs or implementing only the most incrementally cost-effective strategies. It is hoped that the cohort simulation technique in this study – implemented in Microsoft Excel – will be useful and accessible to policymakers and practitioners willing to apply the public health approach, and who can customize it to suit their local context and research needs in an era of widespread global scale-up of treatment and care.

HIV/AIDS prevalence in India: estimates for 2008

Since 1998, for calculating the national HIV prevalence rate in the general population, India has officially relied on sentinel surveillance convenience samples from antenatal clinics (ANCs), sexually transmitted infectious disease clinics (STD), and sites catering to high-risk groups for HIV transmission: injecting drug users (IDUs), female sex workers (FSWs) and men-who-have-sex-with-men (MSMs). The prevalence rate in the ANCs is assumed to generalize to the general adult population of heterosexual men and women, with certain adjustments. This method had previously been the basis of high estimates of PLWHA: 5.2 million HIV+ for 2006 [10]. Based on the findings of a randomized sample of blood tests in a large district of a southern state considered high-prevalence, it was suggested that the actual national prevalence was much lower, perhaps one-third the NACO estimate [11]. Responding to these critiques and the preliminary findings of the random blood-sampling of about 102,000 people under the National Family Health Survey-3 (NFHS-3), NACO revised its method and estimate for 2007 to 2.33 million (+/- 20%), using the 2006 surveillance data and back-projected the revisions for previous years (Fig. 1). The revision has occurred in the background of more confidence in the surveillance system with an increased number and dispersion of the sentinel sites from which the data are collected.
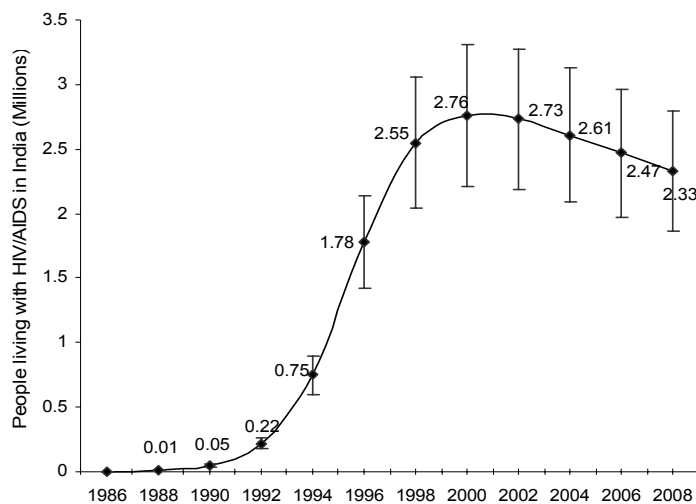
Fig. 1. **People living with HIV/AIDS in India.** Estimates by the National AIDS Council (NACO). Vertical lines show 20% +/- error bars, as reported by NACO. Source: NACO 2007 [3]

However, sources of bias remain, nine of which are noted by NACO [3]. Of these the following two are important, with the first partially incorporated in a re-estimation study of

state HIV prevalence [12]. First, all the surveillance sites are located in government clinics, and hence miss those who seek care from the private sector. Therefore, the ANC prevalence rate (ANC%) may overstate or understate the true rate in pregnant women. Accounting for this bias involves knowledge of the HIV-related characteristics of groups that frequent public sector clinics vs. those who prefer private (Table 1). Second, only 25% of the surveillance sites are rural; mostly located in district centers and concentrated in certain states. If the epidemic in a state has a rural dimension, then the ANC% would lead to an underestimate. These biases are jointly accounted for in Table 1 (see notes). Separately, the availability of population-level prevalence data from the NFHS-3 is utilized to further reduce the bias in generalizing from ANC%. The weighted average of the adjustment multiplier by a state's bracket in Table 1, and the ratio of its NFHS-3 prevalence to 2006 ANC%, is taken as the adjustment factor (background parameters in the appendix). The NFHS-3 ratio is weighted higher for the 7 states that were oversampled and lower for others. Table 2 presents the adjusted total population prevalence, after adding the HIV+ in IDU, FSW, and MSM groups using the 2006 surveillance prevalence rates specific to them.

Table 1. Adjustments to ANC prevalence to act as a proxy for general population prevalence

If:	*And:*		
	If the epidemic is primarily urban	If the epidemic is rural-urban	If the epidemic is primarily rural
On average, public sector patients' SLI is lower than the local median SLI	**I.** § GP% << ANC%	**II.** ¶ GP% < ANC%	**III.** GP% ≈ ANC%
On average, public sector patients' SLI is the median local SLI	**IV.** § GP% < ANC%	**V.** GP% ≈ ANC%	**VI.** † GP% > ANC%
On average, public sector patients' SLI is higher than the local median SLI	**VII.** GP% ≈ ANC%	**VIII.** ¶ GP% > ANC%	**IX.** † GP% >> ANC%

SLI, Standard of Living Index of public sector facility users from Reproductive & Child Health Household Level Survey 2002-2004 [12]; GP%, general population HIV prevalence rate, not including high-risk groups (% of adult population, 15-49); ANC%, HIV prevalence rate in women attending public sector antenatal clinics (% of attendees tested). Roman numerals denote the adjustment brackets subsequently reported in Table 2.

Notes: Whether an epidemic is urban, rural-urban, or rural, is determined on the basis of the 2006 ANC% in sentinel surveillance sites by location [13]. For example, if the mean ANC% in rural sites is substantially higher, the epidemic is rural.

§ In general, public sector facilities receive HIV+ antenatal patients turned away from private facilities on discovery of their sero-status [12]. This phenomenon increases in urban areas. When the epidemic is primarily urban, the effect is an over-concentration of HIV+ mothers at the predominantly urban/periurban sentinel surveillance sites (75% of the total sites in 2006), and the ANC% overstates the true population prevalence rate. This overstatement is more severe if the observed public sector facility SLI is lower than the median local SLI, which implies a local availability of private antenatal clinics and their preferential use by higher income strata.

† Most public sector antenatal or tertiary care facilities are in district centers or urban/periurban areas. This does not restrict rural people from appearing for antenatal care, but if the epidemic is rural and the observed public sector SES is equal to or higher than the median local SLI, then this may indicate that a sample of rural women is lost (e.g., facilities too far for some rural populations), such that ANC% underestimates GP%.

¶ When the epidemic is mixed rural-urban, and if the observed public sector antenatal patient SLI is lower, then ANC% slightly overstates the GP%. This is because the factor of low-income patients turned away from private facilities is still in effect, but compensated by some rural HIV+ mothers lost to sample. In contrast, if the public sector SLI is higher, then the factor of rural HIV+ mothers lost predominates, and GP% may be higher.

Table 2. HIV/AIDS in Indian states, 2008: total prevalence in adult population, 15-49

State	Adj. bracket	Adult Pop. 2008 ('000)	NFHS-3 adult prev. 2005-06 [†]	NACO 2006		Adj. factor[¶]	Adjusted estimates, 2008		
				ANC%	Total adult prev.		GP%	Total adult prev.	Total PLWHA
NFHS-3 oversampled states [§]									
Andhra Pradesh	II	45,883	0.97% (0.77%)	1.35%	1.05%	0.80	1.07%	1.13%	517,230
Karnataka	II	32,055	0.69% (0.68%)	1.02%	0.81%	0.77	0.78%	0.87%	278,715
Maharashtra	II	59,493	0.62% (0.72%)	0.80%	0.74%	0.83	0.67%	0.76%	451,863
Manipur	VIII	1,336	1.14% (1.13%)	1.19%	1.67%	0.98	1.17%	1.63%	21,798
Nagaland	V	1,235	-	1.03%	1.26%	1.00	1.03%	1.13%	13,946
Tamil Nadu	II	37,152	0.34% (0.31%)	0.45%	0.39%	0.82	0.37%	0.40%	150,307
Uttar Pradesh	V	93,605	0.07% (0.08%)	0.19%	0.11%	0.58	0.11%	0.12%	111,998
All other states									
A & N Islands	IV	258	-	0.56%	0.37%	0.95	0.53%	0.54%	1,387
Arunachal Pradesh	IX	677	0.00%	0.08%	0.05%	0.77	0.06%	0.07%	468
Assam	VI	15,954	0.25%	0.04%	0.03%	2.80	0.11%	0.11%	18,171
Bihar	V	45,599	0.00%	0.35%	0.16%	0.67	0.23%	0.25%	116,029
Chandigarh	I	756	-	0.23%	0.34%	0.85	0.20%	0.22%	1,653
Chhattisgarh	V	11,984	0.08%	0.30%	0.17%	0.76	0.23%	0.27%	32,892
Delhi	I	10,122	0.11%	0.18%	0.27%	0.77	0.14%	0.16%	16,182
Goa	II	879	0.44%	0.91%	0.73%	0.79	0.72%	0.78%	6,818
Gujarat	II	31,258	0.17%	0.58%	0.43%	0.73	0.42%	0.43%	134,367
Haryana	VI	13,222	0.10%	0.14%	0.10%	0.94	0.13%	0.14%	18,172
Himachal Pradesh	VI	3,634	0.30%	0.05%	0.03%	2.73	0.14%	0.15%	5,297
Jammu & Kashmir	IV	6,169	0.27%	0.04%	0.04%	2.92	0.12%	0.13%	7,744
Jharkhand	V	15,513	0.11%	0.18%	0.11%	0.87	0.16%	0.16%	24,394
Kerala	III	18,625	0.00%	0.21%	0.13%	0.67	0.14%	0.15%	27,629
Madhya Pradesh	V	35,331	0.37%	0.17%	0.11%	1.39	0.24%	0.24%	86,223
Meghalaya	VII	1,429	0.00%	0.09%	0.06%	0.67	0.06%	0.09%	1,222
Mizoram	VIII	548	0.92%	0.96%	0.74%	1.02	0.98%	1.15%	6,283
Orissa	V	21,493	0.21%	0.43%	0.22%	0.83	0.36%	0.37%	78,785
Pondicherry	V	691	-	0.63%	0.55%	1.00	0.63%	0.66%	4,534
Punjab	II	14,857	0.30%	0.12%	0.12%	1.46	0.18%	0.18%	26,502
Rajasthan	V	32,525	0.08%	0.30%	0.17%	0.75	0.23%	0.23%	76,078
Sikkim	IX	334	0.00%	0.13%	0.08%	0.77	0.10%	0.11%	371
Tripura	VII	1,972	0.00%	0.21%	0.12%	0.67	0.14%	0.16%	3,075
Uttaranchal	IV	4,965	0.00%	0.11%	0.08%	0.63	0.07%	0.08%	3,911
West Bengal	I	48,369	0.18%	0.44%	0.30%	0.70	0.31%	0.35%	167,751
INDIA		607,923			0.36%		0.36%	0.40%	2,411,796

Adj., adjustment; Pop., population; Prev., HIV prevalence (%); NFHS-3, Third National Family Health Survey; Total adult prev., total HIV prevalence in the population, equal to GP% plus prevalence in high-risk groups: female sex workers, injecting drug users, and men who have sex with men. *Notes*: 'Adj. bracket' (Roman numerals) assigns the state to an adjustment category based on Table 2 for generalizing ANC% to GP%. Adult population (2008) estimates are based on Census of India (2001).

[§] The NFHS-3 sampling methodology assigned greater importance to the estimate in the six 'high prevalence states' (all oversampled excluding Uttar Pradesh), and hence increased the number tested. Later, Uttar Pradesh was added as another oversampled state based on its emerging importance as a new epidemic area. The proportion of samples tested over 2005-06 in these seven states under NFHS-3 was 75% of the national total of samples [14].

[†] For NFHS-3 oversampled states, the 15-49 prevalence rate was released (15-54 rate in brackets). For all other states, the 15-49 rate has not yet been released. Therefore, the 15-54 rate is shown, as estimated directly from the survey data files [15].

[¶] The adjustment factor is calculated based on the formula discussed in the text. The data used is reported in an appendix.

Table 2 attests that there are approximately 2.41 million PLWHA in India at the beginning of 2008, or 0.4% of the adult population aged 15-49, within the range 0.27%-0.47% stated for 2007 prevalence [3]. This represents the best available precision after triangulating the data from sentinel surveillance, the national population-based survey during NFHS-3, and other data such as behavioral and socioeconomic surveys. The majority of these PLWHA live in the first six states from Table 2 that are considered high-prevalence in India. Because of a recent decline in new incidence in some of these states, as marked by seroprevalence studies [16], the overall HIV disease profile has matured, and the proportion of HIV+ persons with advanced clinical stages of the disease can be expected to be larger. Fig. 2 illustrates this point with prevalence curves estimated using the revised NACO back-projections [3, 13]. Prevalence rates in 'Old-' and 'Medium-aged' epidemic states have arrived at a plateau at very different levels, or are declining as population grows alongside HIV/AIDS-related mortality. Prevalence rates in 'hotspots', or new epidemic states, is rising, and more individuals here are in earlier stages of HIV, having been infected more recently. The classification of the states by their age of epidemic (Table 3) and the prior calculation of the number of PLWHA in each state are crucial inputs into the calculations that follow.
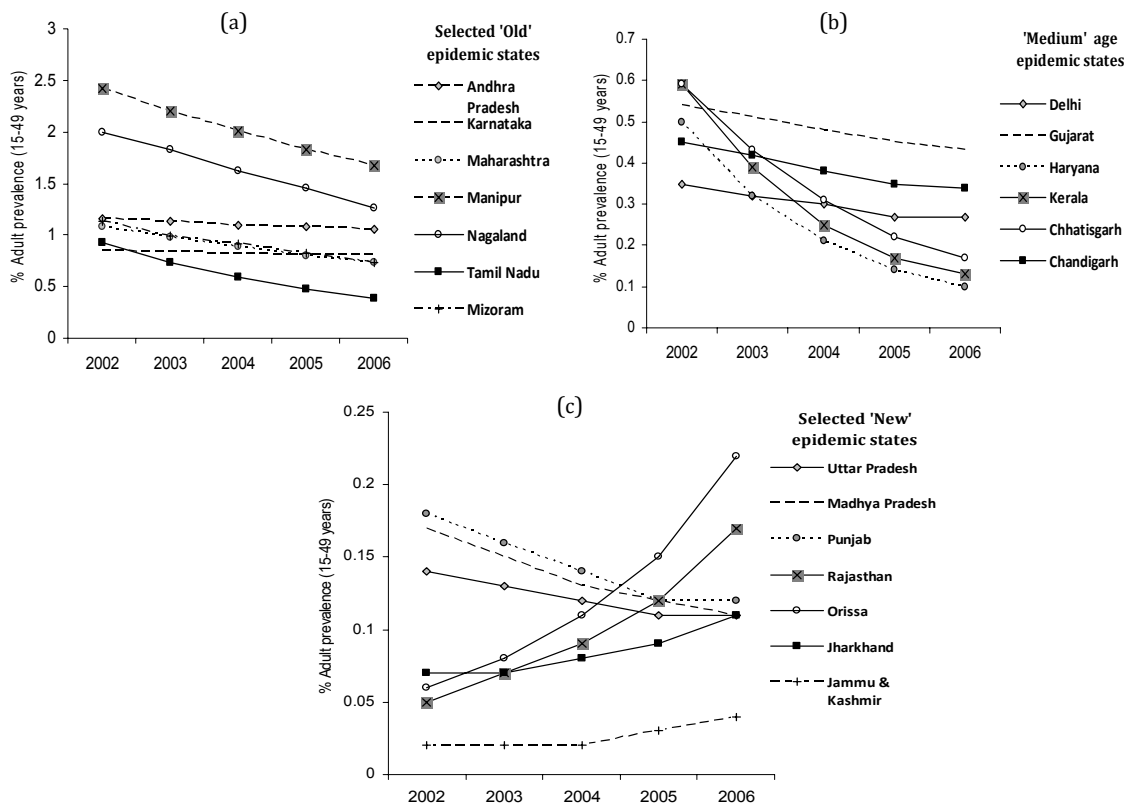
Fig. 2. **HIV prevalence across time in states by epidemic age classes**. Data from NACO 2007 [13]

Table 3. Classification of Indian states by age of local HIV epidemic

Old	Medium	New
Andhra Pradesh	Chandigarh	A & N Islands
Goa	Chhattisgarh	Arunachal Pradesh
Karnataka	Delhi	Assam
Maharashtra	Gujarat	Bihar
Manipur	Haryana	Himachal Pradesh
Mizoram	Kerala	Jammu & Kashmir
Nagaland		Jharkhand
Pondicherry		Madhya Pradesh
Tamil Nadu		Meghalaya
West Bengal		Orissa
		Punjab
		Rajasthan
		Sikkim
		Tripura
		Uttar Pradesh
		Uttaranchal

Source: author calculations, and Figs. 2 (a)-(c)

Demand for treatment in 2008: ART and opportunistic infection prophylaxis

Globally, the number of individuals with HIV/AIDS who received treatment with antiretrovirals has risen dramatically in recent years. In one authoritative study published in the journal *AIDS*, this number was estimated to be about 2.7 (3.38) million by the end of 2007 (2008), with India contributing 4.7% of the total, if its estimated share from 2006 remained stable [17]. However, this share indicates the number of people on ART by mid-2008 of about 126,430-150,000, which seems low considering that nearly 100,000 people already receive free/subsidized ARV regimens (Table 5). If even a small percentage of the remaining 2.41 million PLWHA in 2008 purchase ARVs with out-of-pocket funds, the demand for ART in the country would be well above 150,000.

Estimating the actual total utilization in India, and the unmet demand for both ART as well as a beneficial but inexpensive treatment – prophylaxis for opportunistic infections (OI) with a drug such as co-trimoxazole – requires estimating the size of two groups: those who seek free/subsidized treatment and those paying out-of-pocket (OOP). In India, public sector facilities are usually used by the lowest income quintiles (especially in urban areas), while quality-of-care conscious middle-class and higher individuals utilize private sector care. This tendency is strengthened by several factors. First is the actual constraint that capacity and quality of care in the public sector for acute conditions and those requiring

complex care such as HIV case management, remains very low – the public system accounts for only 20% of all curative care [18]. This results in forcing a form of economic triage that requires any individuals with ability-to-pay to transfer to private providers. Second, low income groups who do present for private care are often redirected to public facilities, a dynamic which is heightened for HIV/AIDS [12]. Third, for HIV+ individuals who can afford to consider it, stigma remains an issue in seeking related treatment. Though private sector costs are significantly higher than in the free/subsidized treatment delivery program, many individuals prefer the privacy of a dedicated personal physician and the choice of not entering public records as a recipient of ART. This preference is more affordable given the price of first-line ARVs and diagnostics has declined even in the private market. Based on these arguments, the percentage of HIV+ individuals in the lowest income quintile (Table 4, estimated from the NFHS-3 sample values), is assumed to be seeking free or subsidized ART ('subsidy-seeking'), and the rest are OOP payers. The proportion 15.9% of PLWHA in the subsidy-seeking arm from the NFHS-3 is adjusted upwards or downwards based on the adjustment brackets previously used. The rest, 84% of PLWHA, are assumed to be preferentially OOP. Table 4 lists the model input values used for baseline estimates.

Table 4. Demand model input parameters

Variable	Value	Reference
Distribution of PLWHA by income		
Population in NFHS-3 sample		
Lowest quintile (subsidy-seeking)	15.9%	
Second quintile	16.7%	[14]
Middle quintile	20.6%	
Fourth quintile & higher	46.8%	
Public sector treatment facilities		
Lowest quintile (subsidy-seeking)	*58.8%*	
Second quintile	*23.7%*	Author
Middle quintile	*9.0%*	estimates [§]
Fourth quintile & higher	*8.6%*	
Adjustment factors for PLWHA proportion in lowest income quintile		
Rural epidemics (brackets III, VI, IX)	1.1	
Urban epidemics (brackets I, IV, VII)	0.9	Author assumptions
Rural-urban epidemics (brackets II, V, VIII)	1	
ART utilization rates (average across CD4 strata)		
Private sector clinic attendees	35.4%	[19]
Public sector clinic attendees	*9%*	[19]
Opportunistic Infection prophylaxis utilization rates (average across CD4 strata)		
Private sector clinic attendees	63.6%	[19]
Public sector clinic attendees	47.8%	[19]

Italicized values are for information purposes only. They were not used in model calculations.
[§] Estimated from six-year patient history data (N: 63,817) received from GHTM, or the Government Hospital of Thoracic Medicine, Chennai (Tamil Nadu). The hospital is a large tertiary institution serving 30,000 HIV+ low-income patients from the southern states: Tamil Nadu, Pondicherry, and Andhra Pradesh. [19].

In the subsidy-seeking group, current utilization is measured from actual numbers in treatment from a variety of official (as of September 2007) and NGO sources. In the OOP

group, utilization is estimated at baseline using the parameter – average across patients in various stages of their disease as would be proxied by CD4 cell counts – from a multi-site, large sample study [19]. This parameter is sensitivity tested further below. After subtracting the current utilization in both subsidy-seeking and OOP, the remaining in the total PLWHA are those receiving no ART. This can be further split into those who receive prophylaxis for OI (using associated utilization rates for subsidy-seeking and OOP from the same source), those who receive no treatment of any kind.

Within the public health approach of an informed use of ARVs, one of the NACO (WHO) guidelines has been to simplify the decision to start ART by establishing markers that are clinical (stage of disease as marked by opportunistic disease or chronic morbidity) and/or virological (using viral loads) and/or immunological (CD4 T-cell count stratum) [20]. In Table 4, the utilization rates from the multi-site study were sample averages across CD4 count strata, and hence are expected to reflect the true population incidence of ART in Indian patients. In general, studies have noted good adherence in India with the NACO starting criteria. However, in subsequent analysis, the eligibility criteria as per WHO guidelines are incorporated to establish the size of cohorts being initiated into ART.

Table 5 reports the current utilization of ART and OI prophylaxis by payment group type (subsidy-seeking vs. OOP), and the sizes of the total population with no treatment at all or with only OI prophylaxis received. Care needs to be exercised in viewing the last column as the total 'unmet' demand, given the need to apply the eligibility criteria. Using a distribution of the HIV+ population in each state across CD4 count strata (based on sources discussed in more detail later) and where states were stratified by epidemic size into different patterns of such distribution, about 1.73 million PLWHA in India would be ART-eligible in the 'no treatment' group by the immunological standard in 2008, of whom approximately 0.82 million currently utilize ART using the rates discussed above (Table 4). This suggests about 0.91 million would be 'NACO-recommended' for ART (CD4 cell count below 200 cells/mm^3), but do not receive it. However, almost all the PLWHA receiving no treatment would 'demand' ART in order to receive health benefits at any stage of their disease – the awareness of treatment availability and its efficacy was very high (98%) in a sample survey (N: 269) weighted towards low-income groups, as reported in one study [21]. Therefore, Table 5 does not engage with clinical/immunological eligibility criteria; these are considered subsequently. The same survey [21] also reported that 90% of the sample was willing to pay OOP for ART.

Table 5. First-line ART and OI prophylaxis utilization and demand across patient types (2008)

	On ART			No ART		No ART, OI prophylaxis Only		No ART, no OI prophylaxis		Prev. rank	ART rank[§]
	Govt. & NGOs[†]	OOP (Est.)	Total (Est.)	Subs. (Est.)	OOP (Est.)	Subs. (Est.)	OOP (Est.)	Subs. (Est.)	OOP (Est.)		
Andhra Pradesh	18,393	154,040	172,433	63,695	281,102	40,526	134,468	23,169	146,634	4	14
Maharashtra	24,047	134,573	158,620	47,667	245,576	30,328	117,474	17,339	128,102	7	7
Karnataka	9,966	83,006	92,972	34,268	151,475	21,803	72,460	12,465	79,015	5	13
Tamil Nadu	19,955	44,764	64,719	3,900	81,688	2,481	39,076	1,419	42,612	11	4
West Bengal	2,019	50,902	52,921	21,942	92,888	13,960	44,434	7,981	48,454	13	21
Gujarat	2,788	40,017	42,805	18,537	73,025	11,794	34,932	6,743	38,093	10	19
Uttar Pradesh	3,510	33,355	36,865	14,265	60,868	9,076	29,117	5,189	31,751	27	16
Bihar	1,089	34,556	35,645	17,326	63,059	11,023	30,165	6,302	32,894	15	26
Madhya Pradesh	1,390	25,679	27,069	12,294	46,860	7,822	22,416	4,472	24,444	16	23
Rajasthan	2,209	22,657	24,866	9,865	41,346	6,277	19,778	3,588	21,568	17	17
Orissa	281	23,463	23,744	12,223	42,818	7,777	20,482	4,446	22,335	12	31
Manipur	3,738	6,492	10,230	0	11,568	0	5,534	0	6,035	1	3
Kerala	2,097	8,073	10,170	2,726	14,732	1,735	7,047	992	7,685	23	5
Chhattisgarh	8	9,796	9,804	5,212	17,876	3,316	8,551	1,896	9,325	14	32
Punjab	1,142	7,893	9,035	3,064	14,403	1,950	6,890	1,115	7,513	19	11
Delhi	3,644	4,910	8,554	0	7,628	0	3,649	0	3,979	20	2
Jharkhand	362	7,265	7,627	3,510	13,258	2,233	6,342	1,277	6,916	21	25
Haryana	594	5,310	5,904	2,578	9,690	1,640	4,635	938	5,054	25	18
Assam	242	5,310	5,552	2,930	9,689	1,864	4,635	1,066	5,054	28	28
Nagaland	511	4,153	4,664	1,702	7,580	1,083	3,626	619	3,954	3	12
Jammu & Kashmir	204	2,350	2,554	902	4,288	574	2,051	328	2,237	26	15
Goa	444	2,031	2,475	638	3,705	406	1,773	232	1,933	6	6
Mizoram	102	1,871	1,973	895	3,415	570	1,634	326	1,781	2	22
Himachal Pradesh	263	1,548	1,811	662	2,824	421	1,351	241	1,473	24	10
Chandigarh	1,224	502	1,726	0	0	0	0	0	0	18	1
Pondicherry	216	1,350	1,566	504	2,464	320	1,179	183	1,285	8	9
Uttaranchal	176	1,187	1,363	383	2,166	243	1,036	139	1,130	31	8
Tripura	4	933	937	435	1,703	277	814	158	888	22	29
A & N Islands	0	421	421	198	768	126	367	72	401	9	30
Meghalaya	4	371	375	171	677	109	324	62	353	30	27
Arunachal Pradesh	12	137	149	70	249	44	119	25	130	32	20
Sikkim	8	108	116	57	198	36	95	21	103	29	24
INDIA	100,642	719,022	819,664	282,618	1,309,514	179,815	626,416	102,803	683,098		

ART, Antiretroviral Therapy; OOP, Out-of-Pocket (patients paying their own expenses); Est., Estimated; Govt., Government of India; NGOs, Nongovernmental Organizations; Subs., Subsidy-seeking patients (lowest income quintiles); OI, Opportunistic Infection; Prev., prevalence (%).

[§] Rank based on value = Number of people on ART (government/NGO – i.e., subsidy-seeking – plus OOP)/Total PLWHA.

[†] Forecast for Q12008 based on actual Government and NGO coverage, September 2007. Data are from National AIDS Control Organization (2007) and NGOs, including YRG Care (Chennai, 2007), Médicines sans Frontières (2008), etc.

The estimates of utilization above imply that the government/NGO sector serves about 4% (5%) of the total PLWHA (eligible PLWHA) with ART in 2008, and 30% of total PLWHA were OOP treatment recipients (38.5% of total eligible by the immunological standard). The proportion on ART was not positively correlated with the prevalence rate in

the state: implying a mismatch that could have economic or infrastructural bases. However, this issue was not investigated further in this study.

Spending on HIV/AIDS treatment in India

With the estimates of utilization of ART, it is straight-forward to calculate the total spending on drugs and associated diagnostics for those in structured therapy. The prices that prevail for the private sector OOP patients (based on a current market survey) are considerably higher than the list prices from Indian generics manufacturers published by the Clinton Health AIDS Initiative as 'ceiling levels' in procurement by the public sector in resource-poor settings, and based on other data on free/subsidized treatment [2, 22]. The non-nucleoside reverse transcriptase inhibitor (NNRTI) based first-line triple drug combination that is commonly prescribed to eligible patients in India without special circumstances is zidovudine, lamivudine and the NNRTI nevirapine, though efavirenz (an NNRTI) is important for the large numbers of patients co-infected with TB.

Table 6. Cost of various ART formulations, by type and data source

Components and strength	Type	Per patient year			Per pill/combination		
		Imunus[1]	Cipla[1]	CHAI[2]	Imunus[1]	Cipla[1]	CHAI[2]
AZT+3TC+EFV (300/150/600)	NNRTI-based first-line combination*	$786	N/A	$540[3]	$2.18	N/A	N/A
AZT+3TC+NVP (300/150/200)	NNRTI-based first-line combination	$337	$404	$174	$0.47	$0.56	$0.24
d4T+3TC+NVP (30/150/200)	NNRTI-based first-line combination	$309	$371	$132	$0.43	$0.52	$0.18
AZT+3TC (300/150)	NRTI-based first-line combination	$322	$387	$129	$0.45	$0.54	
d4t+3TC (30/150)	NRTI-based first-line combination	$155	$192		$0.22	$0.3	
d4t+3TC (40/150)	NRTI-based first-line combination	$165			$0.23		
EFV (600)	NNRTI single drug	$551	N/A	$164	$1.53	N/A	$0.46
NVP (200)	NNRTI single drug	$220	$264	$45	$0.31	$0.37	$0.06
IDV (400)	PI single drug	$792	$950		$0.37	$0.44	
AZT (300)	NRTI single drug				$0.35	$0.42	
3TC (150)	NRTI single drug	$131	$158	$36	$0.18	$0.22	$0.05

NRTI, Nucleoside Reverse Transcriptase Inhibitors: AZT, Zidovudine; 3TC, Lamivudine; d4T, Stavudine
NNRTI, Non-nucleoside Reverse Transcriptase Inhibitors: EFV, Efavirenz; NVP, Nevirapine
PI, Protease Inhibitors: IDV, Indinavir.   CHAI: Clinton Foundation HIV/AIDS Initiative
[1] Based on New Delhi market survey: telephone interviews with private pharmacies (March 2008). [2] Based on latest CHAI price list, May 2007 [23], except [3] based on CHAI (2006) [24].
* Recommended for HIV-TB co-infection. Both Imunus and Cipla are Indian pharmaceutical companies.

The OOP price of the commonest ART combination regimens is twice as high as that paid by the state AIDS control societies and large NGOs, which can place bulk orders with

Indian generics manufacturers and negotiate a lower price [25]. Prices in both OOP and public/NGO arms vary by state and the particular manufacturer (the prices in the OOP market from Cipla were about 20% higher than Imunus). The actual mix of drugs prescribed in India across diverse care settings is difficult to determine; approximate weighted average of prices can be calculated, with prices of the dominant combinations (bold in Table 6) weighted higher. This calculation suggests that for patients on first-line treatment regimens paying OOP, weighted average costs are about US$350 per patient year (ppy) and for the free/subsidized sector, about US$228 ppy at 41 rupees to the US$ (2008).

Using the total utilization figures from Table 5, these price parameters imply that the public sector and NGOs spend about US$23 million p.a. currently on ART drugs alone (not including the cost of OI prophylaxis, diagnostics, chronic care, etc.), 9% of the total of US$274 million spent on first-line ARVs. With the public sector price per CD4 test of about US$18.2 (approximately double this value in the private sector), and an established bi-annual testing strategy, the total spending on CD4-diagnostics is in the range of US$27 p.a. These values were sensitive to the assumed utilization rate in the OOP group on ART (Fig. 3); given it was seven times the size of individuals receiving free/subsidized ART. Viral-load diagnostics, with high reagent and capital costs, remain extremely expensive and were not widely or intensively used in clinical management of patients in India.
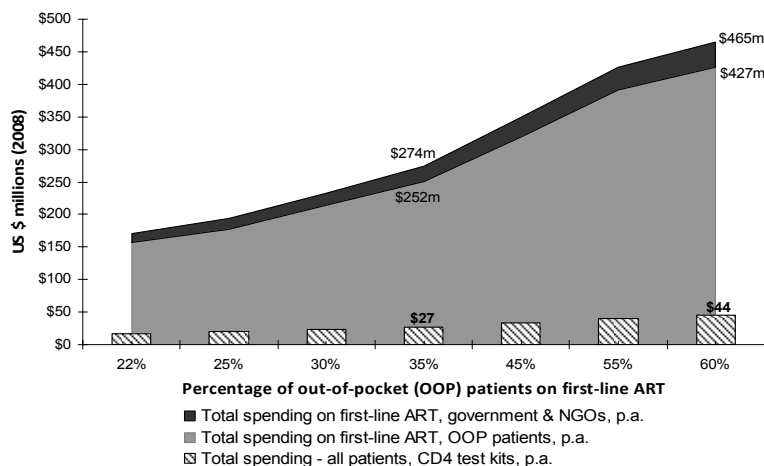
Fig. 3. **Sensitivity of spending on first-line ART and CD4 testing to assumptions on OOP utilization**
Baseline model assumption for percentage of OOP on first-line ART is 35.4%. Assumed average cost of first-line ART for government and NGOs: $228 per patient year (ppy) [2], and for OOP: $350 ppy. Cost of bi-annual CD4 testing: government & NGOs: $18.2 ppy [2], OOP patients: $35 ppy.

Penetration of second-line ARVs remains low in India – the government has very recently announced a plan to initiate free availability of an associated regimen to about

2,000 individuals in 2008, potentially expanding to 5,000 by the end of the year [26]. At NACO's informally declared rates of first-line treatment failure of 3-5%, this initial coverage amounts to 4.8-8% of the estimated group of patients who may need second-line treatment (based on a total ART-recipient population of 0.82 million). The cost of ARVs in a second-line triple-drug regimen comprising a protease inhibitor (PI) matched with two nucleoside reverse transcriptase inhibitors or NRTIs, is high. A study quotes the cost as $1,435 ppy [2]. Indinavir (PI) alone cost $792 ppy, in OOP (Table 6). In an interview, the director of the government Tamil Nadu State AIDS Control Society suggested that costs of the state's limited rollout of second-line ART from January 2008 would be Rs.8,000 ($195) per patient month, or approximately US$2,340 ppy [27]. However, the composition of the 'second-line' regimen is yet unclear, as was the fact whether the costs were of overall patient management in second-line treatment or solely the costs of the drugs, though the discrepancy in case the latter is assumed should be moderate.

Measuring the impact of non-availability of ART for some PLWHA

As estimated above, there are approximately 1.6 million adult PLWHA in India who are not on any ARV program. The question is asked: what will be the deaths and severe illnesses in this group going forward, if none receive ART? The answer has public health and socioeconomic value since the terminal care requirement for patients with near-fatal chronic HIV morbidity as well as curative needs for severe opportunistic disease (OD) episodes impose costs on the health sector; while the families suffer both bereavement as well as a financial burden from the lost income of victims. The disease course and outcomes in this cohort of 1.6 million individuals, almost equally split between those receiving some OI prophylaxis – based on utilization rates seen in India – and those receiving no structured treatment at all, should also illuminate the general case for more aggressive expansion of the government/NGO program for ART – as long as the particular interventions under the program are cost-effective in the Indian setting.

HIV natural history and state-transition simulation

The individuals in the two sub-cohorts: 0.785 million with no treatment at all, and 0.819 million receiving OI prophylaxis using co-trimoxazole as the assumed drug choice, are distributed in different disease stages depending on when they were infected, their

physiology, and any other individual-level variations. As individuals move through progressive disease stages (Fig. 4) their risks of chronic mortality – due to advanced HIV/AIDS related neoplasms and metabolic complications – and mortality due to severe OD increase, given some individual patient variation. At a population level these variations would edge towards average risks seen to occur by various disease stages, where disease stage is established by a viable marker that can and has been recorded consistently for large numbers of individuals. The use of CD4 T-cell count as a marker for the progress of untreated HIV is shown in Fig. 4. The decline in CD4 count – attended in the background by a generalized increase in viral load (viremia) – opens the door for immune suppression and the onslaught of opportunistic infections. These well-known associations of progressive disease with progressive risk of mortality and morbidity are the basis for state-transition simulations [2, 5, 6], whether they simulate individuals (using a stochastic process to seed disease and mortality-related events), or groups of individuals (using average risk).

HIV RNA copies/ml *		CD4 cells/mm3 **
>100000	0.352	<50
30001-100000	0.26	51-200
10001-30000	0.172	
3001-10000	0.103	201-350
501-3000	0.022	
21-500	0.062	351-500
0-20	0.029	>500
Total	1	

* HIV RNA copies/ml distribution at baseline - Source: Freedberg *et al.*, (N=6918) [2]
** Correspondence of HIV RNA copies/ml and CD4 cells/mm3 – Source: author assumptions
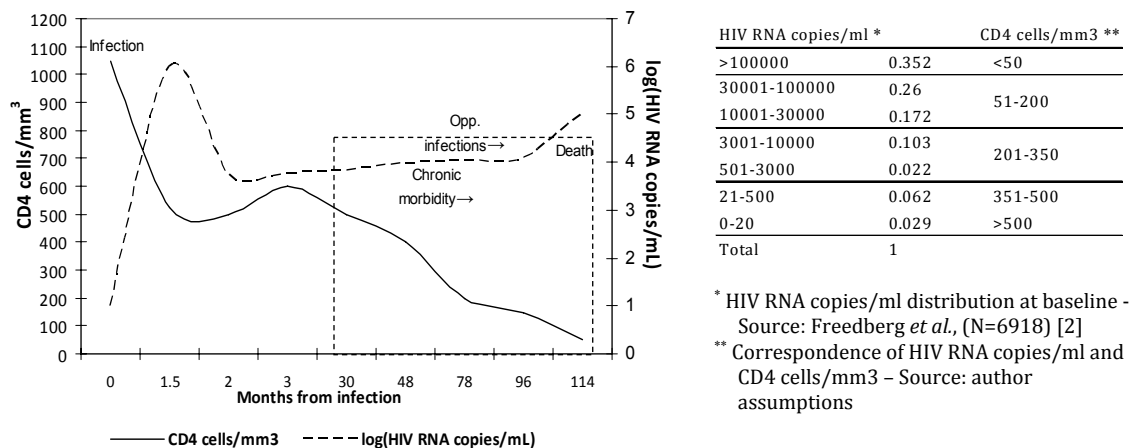
Fig. 4. **Natural history of HIV/AIDS in a hypothetical Indian patient without treatment: correspondence of CD4 count with viral load.** Area within the dashed-line border indicates the zone of interest for simulations in this study: corresponding to CD4 T-cell count < 500 cells/mm³ and viral load > 5000 copies/ml. The dummy data used to construct the graph incorporates a correlation of about -0.6 between viral load and CD4 count in a period after two months since infection. Clinical data in Mellors *et al.* [28], suggests a correlation of -0.42. Adapted from published studies [29] [30] for illustration purposes only. Actual course of disease varies across individuals.

Differentiating states by age of local epidemic

In the absence of India-specific parameters, risks of chronic mortality per month in HIV+ patients by three cumulative CD4 strata (Table 7) are available from a clinical trial in Côte d'Ivoire, and were previously used in state-transition models [2, 5]. However, India-specific risk of opportunistic infections were available by a more differentiated set of six

CD4 strata from a large-N, previous observational study in southern India, also recently used in simulation [2]. With the intention to exploit such differentiated data, the cohort of 1.6 million individuals to be simulated was segmented into seven CD4 strata (Fig. 5). The stratum 201-350 cells/mm3 was split in order to specify certain ART starting criteria in later analysis. As previously discussed, with large enough numbers – as obtains in India – the disease stage pattern of individuals among a states' PLWHA is related to the overall age of the epidemic (assuming that the HIV+ infected do not move, or move back to home states on progress of disease to seek support from caregivers). Using previously discussed trends in states' incidence and prevalence (Fig. 2), observations on HIV disease natural history (Fig. 4) in India [29, 31], as well as some large-N studies of CD4 proportions of HIV+ patients at baseline in different states representing different epidemic types [32-34], the distribution patterns into CD4 strata by epidemic age were assumed as per Fig. 5 below.

CD4 cells/mm3	By epidemic age			Other sources		
	Old	Medium	New	Rajasekaran et al. (N=1370)	Cecelia et al. (N=3104)	Ghate et al. (N=62)
0-50	0.24	0.190	0.110	0.277	0.13	
51-100	0.25	0.205	0.150	0.244	0.32	0.85
101-200	0.325	0.300	0.190	0.328		
201-250	0.04	0.075	0.095		0.21	0.06
251-350	0.06	0.100	0.150	0.151		
351-500	0.05	0.080	0.185		0.13	0.08
>500	0.035	0.050	0.120		0.21	
Total	1.0	1.0	1.0	1.0	1.0	1.0
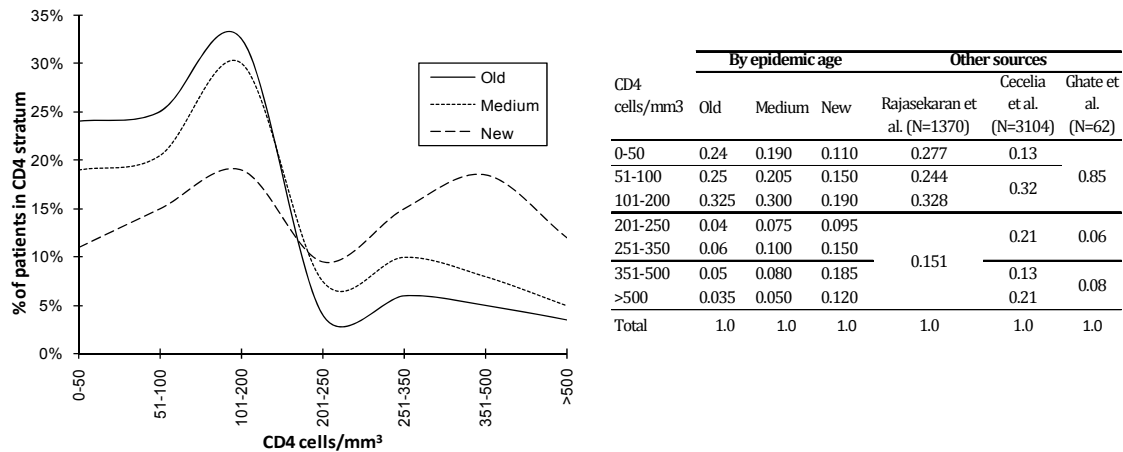
Fig. 5. **Expected average patient mix at baseline by stage of disease as proxied by CD4 count, across epidemics of varying age.** The shapes of the curves are also suggested by the prevalence curves over time in Figs. 2(a)-(c). The 'bump' at the end of the curves (higher CD4 counts) accommodates recent incidence, higher for new epidemic states (more patients with recent infection histories). Other sources: [32-34]

The individuals in each of the two groups ('sub-cohorts') within the cohort of 1.6 million – no treatment at all, and OI prophylaxis only – were estimated previously, by state. They are now distributed to seven CD4 strata as in the table above (Fig. 5). Given an appropriate state-transition method to simulate disease progress and the assumed risks of monthly opportunistic disease and mortality per stratum as discussed, the disease course in the two sub-cohorts can now be simulated for any desired length of time, in a model which would match both epidemiological and clinical contexts as seen in India (within a feasible modeling paradigm). Outcomes would be deaths and severe opportunistic disease episodes.

Cohort simulation for the total population of PLWHA without ART

The model utilized here is a state-transition model that moves groups in different CD4 strata – rather than individuals as in other, recent state-transition simulations – as per the likely disease course in the absence of antiretroviral treatment. At the beginning of the simulation – assumed to be February 2008 – all individuals, by Indian state, in each sub-cohort, are allocated to specific CD4 strata as per methods already discussed. Individuals move from higher CD4 counts to lower, proxying a decline in their immune health and a background worsening of the viral load. The rate of movements of the individuals in the groups within each stratum per month are shown in Table 7 below, and are modeled after published studies of the natural history of HIV disease in India [29, 35, 36] to capture the observed quickening of the pace of immune decline in individuals with more advanced disease, and the relative width of the stratum in terms of cell count range.

The rates of inter-strata transfers in Table 7 lead to a depletion of each stratum at 6 months and 12 months – without additions from a higher cell count stratum – that mimics the mean monthly cell count decline rates by viral load strata in a published study [2] themselves based on data from a large-N observational database [33]. The association of CD4 count strata and viral load strata as per Fig. 4 was used. As an example, that study implies a mean monthly decline rate of 5.4 cells/mm^3 in stratum (3) from Table 7. If the median cell count in this stratum is 174 cells/mm^3 (based on how many standard deviations away the median – rather than mean – in the stratum is from the overall sample mean across strata in the large N database of 318 cells/mm^3, σ: ±291), this implies that by month 12, 50% of those in the stratum should transfer to the next lower stratum: (2). The monthly stratum decline rate of 7% roughly achieves the same effect, even after allowing for some error in estimating the median cell count in that stratum.

Table 7. Simulation model input parameters

CD4 cells/mm^3 strata	Baseline monthly combined risk of severe opportunistic infections §		Reference
	No ART, no OI Prophylaxis	Co-trimoxazole OI prophylaxis	
(1) 0-50	14.66%	5.71%	
(2) 51-100	9.85%	3.20%	
(3) 101-200	5.27%	1.92%	
(4) 201- 250	2.60%	1.12%	[2]
(5) 251-350			
(6) 351-500	1.14%	0.43%	
(7) >500	0.84%	0.44%	

Table 7 continued

	Baseline monthly risk of chronic mortality		
	Without prior OD	With prior OD	
0-50	8.99%	7.69%	[2]
51-200	0.91%	4.48%	
>200	0.33%	0.66%	
Transferring CD4 count strata	Monthly inter-strata transfers under no ART		
(2) to (1)	10%		
(3) to (2)	8.5%		Assumptions on rate of disease progress based on [2, 29, 30, 35, 36]
(4) to (3)	7%		
(5) to (4)	5%		
(6) to (5)	5%		
(7) to (6)	5%		

OD, Opportunistic Disease; OI, Opportunistic Infection.

§ The risk in this table is a composite of the individual monthly risk (%) for the following severe opportunistic infections, risk for CD4 strata (4) & (5) shown in brackets: bacterial (0.02%), fungal (0.08%), tuberculosis (1.08%), malaria (0.02%), toxoplasmosis (0.08%), *pneumocystis jiroveci* pneumonia (0.08%), other severe OI (1.24%). Data is drawn from Freedberg *et al.* [2], who in turn derive it from an observational database of patients in Southern India [33].

The simulation was run for five years. Results are reported in Table 8 and Fig. 6. The five-year mortality rate is high at 84.74% in the cohort of 1.6 million (rates in each sub-cohorts are shown below). Though the cost of terminal care was not estimated, the overall expense in treating severe opportunistic disease (OD), was very high, above US$280 million in the cohort, even with a conservative cost of such treatment calculated from previously published, disease-specific cost parameters [2]. Treatment with co-trimoxazole was a good return on investment and reduced OD morbidity, though not overall mortality given that no direct effect on HIV disease progress or chronic mortality risk was modeled (there is no broad support for such linkage). These results echo previous findings on co-trimoxazole cost-effectiveness in sub-Saharan Africa [37]. The Government of India should accelerate the wide usage of OI prophylaxis, while using a wide suite of such drugs to prevent the emergence of resistance to co-trimoxazole, especially in bacterial pathogens.

Table 8. Results of a simulation of disease course for a cohort of patients receiving no ART

Variable (all costs in 2008 US$)	No treatment, no OI prophylaxis	Co-trimoxazole OI prophylaxis only
Number of HIV+ individuals	785,898	806,234
Length of simulation	60 months (Feb. 2009 – Feb. 2014)	
Total mortality	667,292	681,816
Total mortality rate	84.91%	84.57%
Number of severe OI episodes	1,689 million	0.656 million
Total 60 months cost of treating OI episodes [1]	$202.8 million	$78.7 million
Total 60 months cost of co-trimoxazole OI prophylaxis [2]	-	$20.9 million

[1] Assumed treatment cost of an episode of severe OD = US$120, as a weighted average of the cost for different severe OI, where the weights are the respective monthly risks of infection. Costs and risk from Freedberg *et al.* [2]. [2] Cost of co-trimoxazole prophylaxis for HIV+ patients, in public sector/NGO facilities was $9.29 ppy [2]. It is assumed that the cost for OOP patients is 30-40% higher, therefore the average price used in these calculations for all patients is $11.4 ppy, or 20% higher than for the subsidized patients used in later analysis.
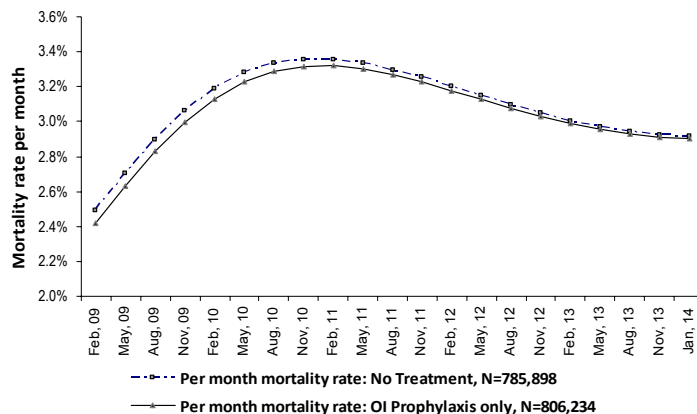
Fig. 6. Simulated mortality in a cohort of 1.6 million HIV+ individuals with no ART

Methods: measurement of the cost-effectiveness of various treatment options

The estimates of spending in the public sector on ART indicate that as a percentage of the total spending, it is still very low. Historically, spending on prevention, care, and counseling have dominated the HIV/AIDS budgets in the country, large by Indian standards, primarily financed with grants and subsidized loans from international partners. With the ongoing sero-stabilization and with the strategic interest in increasing survival rates and structured management of HIV, the question is not whether ART must be scaled up – but with what treatment models and to what level of coverage. Any chosen level of increase in government financed ART must meet public finance criteria. At the minimum, this means particular treatment models must demonstrate incremental cost-effectiveness. Previously, the potential epidemiological effects on transmission of increased treatment (via disinhibition of risky sexual intercourse, etc.), and budget trade-offs vis-à-vis prevention were also included in this debate [38]. With the notable prevention success of recent years, the benefits in avoiding premature mortality and excess morbidity in treatment programs are the more pertinent criteria for evaluation. Separately, studies have indicated that the OOP cost of drugs and diagnostics is one of the significant predictors of failure to adhere to the drug regimen in patients on self-funded ART [39-41], which in itself leads to viral resistance in the individual and the increasing risk of circulation of resistant viral strains in the population. If this effect is considered, free or subsidized ART which involves government-approved, structured care (i.e., drug regimens and clinical decision-making based on guidelines, and monitoring of adherence) would confer additional benefits.

151

However, these benefits are outside the scope of this study, which focuses on cost-effectiveness estimated on the basis of the years of life saved in treatment, and the averted severe morbidity.

Simulated ART strategies

The state-transition model based on CD4 count strata as used above for the case of a cohort of 1.6 million PLWHA without ART is employed again for simulating policy-relevant treatment models and their specific variations based on initiation (staring) criteria. Efficacy of ART is based on parameters in Table 9 and is discussed further below. The total size of the simulation cohort was 122,947 based on approximately 15% of the 0.785 million PLWHA in 2008 estimated previously, who received neither ART nor co-trimoxazole OI prophylaxis. It also represents an approximate doubling of the size of public sector free ART program, assuming that this program (currently at 100,000 individuals) grows to 120,000 by the end of 2008. The cohort was assumed to be recruited in a treatment-naïve status, and begun on ART therapy en masse in February 2009.

Two treatment models involving 'single-line' ART regimens – patterned on the widely prescribed first-line combination of an NNRTI plus two NRTI drugs – were simulated for a period of five years, assumed to be the maximum length for which still-alive patients, follow up with the treatment provider and maintain the high adherence (95% or above) required to model effective ART. This is still generous, considering one experience of 60% loss-to-follow-up (failing to report for at least one visit in the second year since starting ART) in India [40], and generally in resource-poor settings. For example, the proportion of those lost to follow-up was as high as 15% in an international cohort study (N: 4,810) [1].

A two-line treatment regimen – first-line therapy as above, followed by a second-line therapy on its failure – was patterned on a commonly prescribed combination in India for such contexts: a PI plus two NRTI drugs. This treatment model was simulated for a period of seven years, assumed to be a longer length of realistic treatment follow-up than single-line therapy: given the cost of the two-line regimen, the enhanced involvement of patient and provider correlated with the cost and complexity of the treatment, and the increased life-expectancy based on prior studies of a two-line regimen. Some further issues related to 'follow-up' as a realistic treatment horizon are addressed in the results section.

The two treatment models involving a single-line ART regimen (no availability of second-line drugs) vary on how treatment failure is handled by the managing physicians. In

one model, 'first-line till end-of-follow-up', though a form of treatment failure occurs, physicians do not stop ART, and continue treatment till death or end of the simulated period (five years). In the other treatment model, physicians determine treatment failure based on guidelines, and stop single-line ART. Those for whom treatment is stopped move to an arm of the simulation where they receive OI prophylaxis, but their disease begin to progress as per a no-treatment cohort (using the state-transitions previously described in Table 7). The rates at which physician-determined treatment failures occur per month are CD4-count-stratified (see Tables 10-11). The rates are based on the likelihood of the occurrence of several factors associated with the NACO guidelines on defining failure of first-line ART.

In total, there are three treatment models – two are single-line, and one is a two-line strategy. Each has two variations related to ART-initiation criteria at baseline. The choice of starting criteria can be based on CD4 cell counts, i.e., starting at a count below 200 cells/mm^3, i.e., with sicker individuals; or starting earlier in the disease stage of a potential cohort, with counts below 250 cells/mm^3. These thresholds have been publicized by WHO under its public health approach to ART, and adopted by NACO [20, 42]. At present, NACO suggests definitely treating if CD4 testing is available and the count is below 200 cells/mm3, and recommends treatment at higher counts if the patient is in an eligible WHO clinical stage (determined on the basis of severity and type of HIV-related clinical disease). Since CD4 testing at baseline is assumed for this simulation, the two variations for each single or two-line model are starting with a cohort with CD4 below 200 cells/mm^3 (later), or below 250 cells/mm^3 (earlier). These variations are also evaluated in similar studies [2, 5, 6].

Baseline assumptions for ART efficacy

All modeled treatment strategies include co-trimoxazole OI prophylaxis and therefore enjoy the reduced monthly risk of opportunistic infection, as previously discussed (Table 7). In the absence of ART, disease progress occurs as per the state-transition rates by CD4 strata presented earlier for a no-treatment cohort simulation. Generally, the effect of ART in responsive patients with regular adherence (at least 95% compliance with the prescribed regimen, as a percentage of doses taken correctly, and on time) is initial, rapid suppression of the viral load and a resultant increase in CD4 cell count. The efficacy of first- and second-line treatment is sourced from Freedberg *et al.* [2] who in turn obtain rates of CD4 cell count increase by 6, 12 months from clinical trials in resource-poor settings.

Using a similar logic as used to determine the state-transition rates by CD4 count strata for disease progress in absence of ART, the effectiveness of ART in this simulation obtains as a movement of certain proportions in each CD4 stratum to the next higher stratum, representing an overall improvement in immune health in the cohort in treatment. Based on the graphs of time distribution of ART effect on CD4 cell count (and viral load) in two simulated, randomly selected patients from a study which also used clinical parameters [6], the immune-improvement response was higher when CD4 cell counts were lower (strata (1) to (3) from Table 7) , and as higher levels of cell counts (stratum (4) and above) were reached, the improvement rate gradually hit a plateau, unless first-line treatment failure occurred (in which case immune retrogression would begin). The efficacy rates assumed in Table 9 were sensitivity tested, and those results are presented further below. An effect independent of the improvement in CD4 counts under ART is the reduction in risks of contracting an opportunistic infection, and of chronic mortality, expressed as multipliers on the CD4-stratified risks noted earlier [43]. Table 9 shows the values sourced from the literature. In sensitivity testing, removal of this independent effect was significant.

Table 9. Model assumptions on efficacy of combination antiretroviral therapy

Variable	Monthly combined risk of severe opportunistic infections §		Reference
Independent effect of combination ART (multipliers on baseline risk)			
On monthly chronic mortality risk	0.24		[5, 43]
On monthly severe OI risk	0.64		
Transferring CD4 count strata	Monthly inter-strata transfers		
	NNRTI-based first-line	PI-based second-line	
(1) to (2)	15% †	16.5%	For first-line combination ART, based on CD4 cell increase rates§ in [2]; and simulated patient history curves in [6]. For second-line, based on [2].
(2) to (3)	12.5%	13.0%	
(3) to (4)	10%	11.5%	
(4) to (5)	7%	8.5%	
(5) to (6)	5%	6.5%	
(6) to (7)	5%	5.0%	
(7)	-		

NNRTI-based combination ART: Includes at least one NNRTI, and two NRTIs. For example: NVP + d4T/3TC
PI-based combination ART: Includes at least one PI (ritonavir-boosted), and two NRTIs, e.g., IND/r + ddl/ABC
§ Mean CD4 cell increase under NNRTI-based ART: 132 cells/mm^3 at 6 months (24 weeks). Mean CD4 cell increase under PI-based ART: 180 cells/mm^3 at 12 months (48 weeks). Source: Freedberg *et al.* [2].
† For NNRTI-based ART, the associated rate of 15% implies that ceteris paribus, after 6 months, 2/3^{rds} of the individuals in the CD4 count stratum (1), i.e. CD4 count 51-100 cells/mm3, would transfer into stratum (2), i.e., CD4 count 101-250, i.e., become immunologically healthier.

Modeling treatment failure

As small-sample observational studies attest, resistance to first-line therapy is becoming increasingly common in India, leading to declining response to treatment. Studies

tracking patients in Mumbai (N:208) found primary resistance rates of 6.7% to reverse transcriptase inhibitors, and a study in Chandigarh (N:60) found 32% primary and 78.3% partial resistance to the NRTIs lamivudine and zidovudine, respectively [44, 45]. However, some additional small sample studies failed to find detectable levels of resistance mutations [46]. The true population prevalence of first-line ARV-related viral resistance remains unknown, which causes NACO officers to suggest that it lies in the range 3-5% at any point for those on such regimens. At what point does such 'declining response' related to resistance equate to treatment failure, such that a decision on continuation, stoppage, or switching of first-line ART is merited? Based on WHO guidelines, NACO has implemented protocols for clinicians to define and act on treatment failure (Table 10). Using only the immunological definition below, a large sample study in southern India (N: 1,370) found a cumulative incidence of 3.9% (95% CI of 2.9 to 4.9) for treatment failure in its cohort [32].

Table 10. NACO guidelines for physician's definition of failure of first-line ART [42]

Clinical failure	New or recurrent WHO stage 4 condition, after at least 6 months of ART [i, ii]
Immunological failure	- Fall of CD4 count to pre-therapy baseline (or below) 50% fall from the on-treatment peak value (if known) - Persistent CD4 levels below 100 cells/mm [ii, iv]
Virological failure	Plasma viral load > 10,000 copies/ml [v]

[i] Certain WHO clinical stage 3 conditions (e.g. pulmonary TB, severe bacterial infections) may indicate treatment failure and thus require second-line therapy to be considered.

[ii] Some WHO clinical stage 4 conditions (lymph node TB, uncomplicated TB pleural disease, oesophageal candidiasis, recurrent bacterial pneumonia) may not indicate treatment failure and thus second-line therapy need not be considered.

[iii] Without any concomitant infection causing transient CD4 cell count decrease.

[iv] Some experts consider persistent CD4 cell counts of below 50/mm^3 after 12 months of ART to be more appropriate.

[v] The optimal viral load value at which ARV drugs should be switched has not been defined. However, values of more than 10,000 copies/ml have been associated with subsequent clinical progression and an appreciable decline in the CD4 cell count.

Physician-determined treatment failure (TF) is a factor in treatment models where first-line therapy is stopped ('single-line ART till TF'), or switched to second line (two-line ART). It is assumed that the guidelines above for immunological failure are used by physicians, which imply that any patients with CD4 below 50 cells/mm^3 after 12 months are 'very likely' to be defined as not responding to treatment and would be either switched to second-line ART or their ART stopped. In higher CD4 count strata, a certain proportion would be at the risk of having reverted to, or below their baseline CD4 count (this risk is considered higher in those in the lower strata after 12 months, after adjustments for the cell count width of the strata). Given that CD4 counts are only measured twice a year, and some terms are not precisely defined, for example, 'persistently below' in Table 10, it is assumed that in a certain proportion of cases in months 12 to 24 where a physician in a developed

country would have determined treatment failure by WHO guidelines, the Indian clinician adopts a 'wait and see' approach. This tendency is considered higher (a smaller multiplier of 0.33 on the rate of determination of TF) when stoppage is the decision, i.e., in the single-line ART till TF model, and lower (a higher multiplier of 0.66) when switching is considered, i.e., a two-line model. This is intuitive when patient-provider interaction is considered, such as the likelihood of slower (than no treatment) progress to morbidity and mortality if first-line ART is continued, even in the presence of resistance and increasing viral load.

Table 11. Model assumptions on different forms of treatment failure

Variable	Value	Reference
Immunological retrogression rate in combination ART, in absence of physician-determined treatment stoppage	4% per CD4 stratum per month after 12 months of treatment [1]	Author assumptions
Physician-determined treatment failure (ending or switching first-line treatment), after 12 months		
CD4 cells/mm³ strata	Baseline rate	
(1) 0-50	80.0%	
(2) 51-100	25.0%	
(3) 101-200	12.5%	Based on
(4) 201- 250	8.5%	[42]
(5) 251-350	7.0%	
(6) 351-500	5.0%	
(7) >500	3.9%	
Physician decisions on treatment failure, Year 2 (months 12th -24th) provisions for CD4>50 cells/mm3		
If treatment options available:	Then multiplier on baseline rate:	Author
No second-line treatment options	0.33	assumptions, physician
Second-line treatment option available	0.66	interviews

[1] This applies to the first-line only treatment models which continue ART till 'end of follow-up' rather than terminating at physician-determined treatment failure (TF). It can be assumed that physician-determined TF, which encompasses virologic, immunological, and clinical markers, would be comprehensive and the rates (80% and below) 'catch' immunological retrogressives as well. The same rate of immunological retrogression is assumed in the PI-based ART period after 12 months of that treatment, applying till end of follow-up (at seven years). Sensitivity analysis follows in Fig. 7 and Table 17 below.

Precisely to respond to the latter intuition, a separate single-line model is considered where clinicians do not stop first-line ART at all ('single line till end-of-follow-up', or EOFU). In the background of this ART model (and in the PI-based arm of the two-line model), the proportion of patients with resistance to treatment will respond less on an immunological basis even if no TF will be diagnosed. The mutations related to resistance accumulate over time naturally because of high replication error and recombination in HIV, as well as because a small fraction of the cohort is assumed to be initially infected with a resistant-mutated viral strain. The rate at which such immunological retrogression occurs is something that has to be assumed in the absence of data. A monthly rate of 4% is used, which is in resonance with the overall 'incidence' or prevalence – depending on the viewpoint – of resistance to first-line drugs (also assumed to apply to second-line drugs

after 12 months of that treatment). The same rate of retrogression is used across the CD4 count strata. In their large sample study, Rajasekaran *et al.* [32] found immunological TF rates to be roughly equal across baseline CD4 counts. The assumed rate of 4% is sensitivity tested, and found to be significant (see Fig. 7, and the sensitivity sub-section).
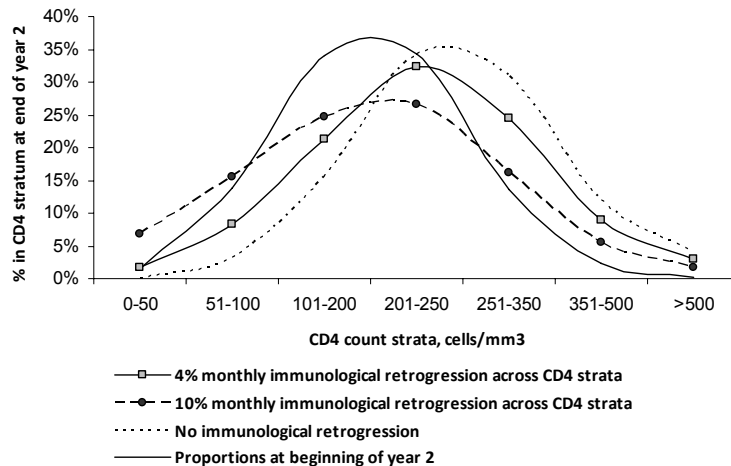
Fig. 7. Impact of resistance to first-line combination ART as immunological retrogression, on the simulated efficacy of combination ART during a year of treatment: case of 'single-line ART till EOFU'. Sensitivity analysis for the assumed rate of monthly retrogression. In the absence of retrogression, ART achieves a shift in the proportion of patients in lower to higher CD4 strata, as shown by the dotted curve (no markers). With 4% monthly retrogression - i.e., 4% of stratum (2) moves to stratum (1), against the direction of ART effect – the curve shifts left. This shift is more marked with 10% monthly retrogression.

Table 12. Model variables: cost

Variable	Costs		Reference
Cost of treatment	annual	monthly	
Co-trimoxazole OI prophylaxis	$9.5	$0.79	[2]
NNRTI-based first-line ART	$225	$18.75	[2]
PI-based second-line ART	$1,435	$119.6	[2]
Weighted average cost of severe OI treatment (per visit)			
Baseline CD4<250 cells/mm3	$119.9		Calculated based on [2]
Baseline CD4<200 cells/mm3	$120.4		
Costs of chronic care*			
CD4 stratum (cells/mm3)	monthly		
(1) <50	$10.25		
(2)-(3) 51–200	$4.1		Based on [2], adjusted to 2008.
(4)-(5) 201–350	$3.75		
(6) 351–500	$3.5		
(7) >500	$3.25		
Terminal care cost (last month of life)	$10.3		[2]
Cost of CD4 cell count, per test	$9.1		[2]

* The original source presented chronic care costs as daily values covering inpatient and outpatient services for routine visits of enrolled patients [2]. For converting daily values to months, it was assumed that, excluding visits to collect antiretrovirals or for medical care following severe OI, the average rate of chronic care visits per month across the cohort was equal to one, and the system incurred a cost per living patient per month as shown in Table 12. Sensitivity analysis on the cost of chronic care was performed and is shown further below, Fig. 9. (c).

Cost of treatment was previously discussed. The values in Table 12 are sourced from a recent state-transition modeling study of cost-effectiveness [2], such that one has comparable results. Some costs have been slightly adjusted to account for inflation. The costs of ART in the public sector have largely remained stable and are not adjusted.

Results: cost-effectiveness of single-line and two-line ART strategies

Six simulations were run for the cohort of 122,947 individuals, corresponding to the three ART models and their two variations based on different CD4-count-driven inclusion criteria. The single-line ART models (till TF and till EOFU) were simulated for five years, and the two-line model ran till seven years. Mortality over time is shown in Fig. 8.
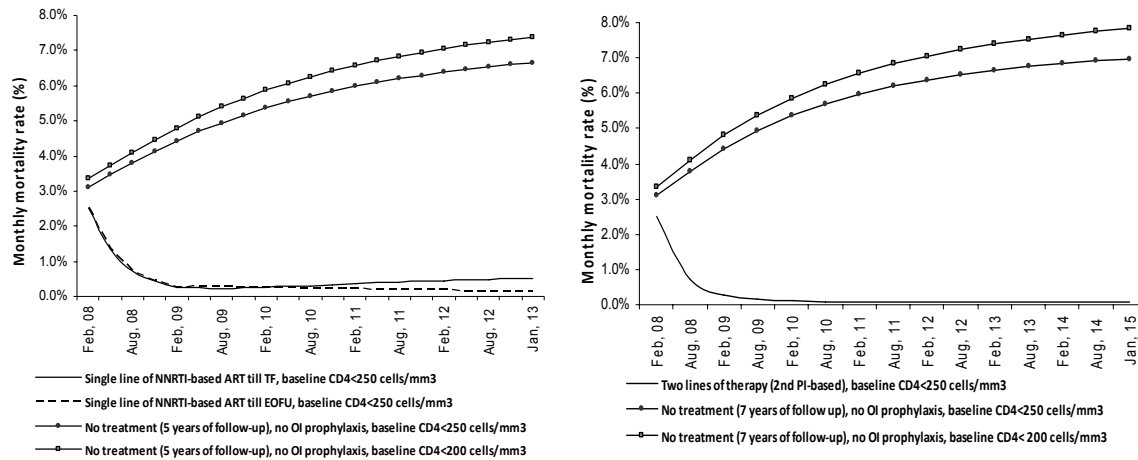
Fig. 8. Mortality in a simulated cohort under different ART models and starting criteria.
TF, Treatment Failure; EOFU, End of Follow-Up period.

Base cases of the simulation result in mortalities, severe OD related morbidity, the overall cost, and the discounted average cost per person over the duration of follow-up for each treatment model and its two variations. Incremental cost-effectiveness analysis against no treatment was conducted using standard methods, utilizing the total cost – which includes expenses related to mortality and morbidity – and the years of life saved (YLS). The incremental cost-effectiveness ratio (ICER) is the ratio of the total cost to the years of life saved and is measured in the unit dollars per life saved. The simulation allows the estimation of quasi-YLS (presented in Table 14), based on the mortality across time in the cohort, and the number alive at the end of the simulation (i.e., at the 61st month, or the 85th). This is not the same as estimating YLS from comparing life expectancy in a no treatment arm vs. that in a treatment arm, since this simulation does not project the future life years

left for still-alive patients at the end of follow-up. For Table 13, the years of life gained in single-line and two-line ART compared to no treatment is from the study by Freedberg *et al*. [2] who assumed a flat treatment horizon of 10 years. The following defense is offered.

First, the duration of follow-up, 5 or 7 years as per the treatment model, is close to the respective undiscounted average life-expectancies with treatment in Freedberg et al. (6-8.6 years). Second, any discrepancy between the two is lessened by discounting future life years in Freedberg *et al*. [2]. Third, as discussed previously, it was important to estimate lives saved only within a follow-up period in which there is a realistic likelihood of maintaining high adherence and visits to the facility. However, there were individuals left alive in the treatment simulations in this study at the end of the duration of follow-up. There are no large sample estimates from India of subsequent survival length in individuals coming off treatment histories with such long durations, and no additional parameters were assumed. Finally, the YLS from Freedberg *et al.* offers the small benefit of incorporating what is the proportionately small risk of non-AIDS mortality, adjusted by sex and age, as estimated from the country's life tables. The values of ICER and the ranking of treatment models and variations are presented in the appendix using the quasi-YLS of this study.

Table 13. Averted deaths and cost-effectiveness of various ART models and starting criteria

ART models and starting criteria (N=122,947)	Years of follow -up[i]	Premature deaths averted compared to 'No ART'	Total costs for cohort, till end of follow-up (US\$ mil.)	Discounted YLS compared to no treatment [ii]	Discounted average per-person total costs till EOFU (2008 US\$)	ICER (US\$/YLS), compared to no treatment	Result[v]
No ART	5	-	\$52.9-53.2[iii]	2.875	\$1,248	-	-
No ART, co-trimoxazole	5	127-169[iii]	\$29.9-31.3[iii]	2.9	\$959	N/C	-
One line of ART, NNRTI-based (till end of follow-up)							
Baseline CD4< 200 cells/mm³	5	93,401	\$164.3	2.33	\$1,450	\$513	Dominated
Baseline CD4< 250 cells/mm³	5	93,535	\$165.7	2.43	\$1,446	\$494	Dominated
One line of ART, NNRTI-based (till treatment failure)							
Baseline CD4< 200 cells/mm³	5	86,294	\$146.4	2.33	\$1,305	\$456	Dominated
Baseline CD4< 250 cells/mm³	5	86,643	\$147.1	2.43	\$1,296	\$437	
Two lines of ART, NNRTI-based, then PI-based[iv]							
Baseline CD4< 200 cells/mm³	7	98,543	\$476.6	4.19	\$6,412	\$1,025	Dominated
Baseline CD4< 250 cells/mm³	7	99,359	\$476.9	4.35	\$6,286	\$979	

EOFU, End of Follow-Up period (5-7 years, depending on treatment model); TF, Treatment Failure; mil., millions; YLS, Years of Life Saved; ICER, Incremental Cost-Effectiveness Ratio = (total cost till EOFU)/YLS; N/C: Not calculated
[i] For all ART or no ART variations, during the years of follow-up, bi-annual CD4 tests, chronic care management, terminal care, and severe opportunistic infection care are offered to all patients and costs incurred.
[ii] YLS based on life expectancies under treatment vs. no treatment in Freedberg *et al.* (2007) are used here. Results using the quasi-YLS values calculated in this simulation, provided in an appendix do not involve major change in the rankings of ICER.
[iii] Depending on baseline CD4 characteristics of recruited cohort (higher values are for baseline CD4<250 cells/mm3)
[iv] Compared to no treatment, no OI prophylaxis, with 7 years of follow-up. This ensures that extending life under a two-line strategy is not penalized in terms of ICER by comparing with a shorter follow-up in the no treatment variation.
[v] All of these were strongly dominated on the basis of average per person costs till EOFU: higher ICER and higher cost.

A detailed discussion of the rankings of the treatment models and starting criteria variations is provided in the concluding 'discussion' section. Some comments are presented here on the averted deaths in the results from Table 13. Though the two-line model runs for two more years of follow-up, it does not avert a significantly higher number of fatalities than single-line till EOFU. The two-lines of ARV strategies incur nearly US$300 million in additional costs compared to even the more expensive single-line ART till EOFU. This implies the incremental cost-effectiveness of extending a second-line vs. ending first-line ART at treatment failure should be of some policy interest (as presented in Table 15).

Table 14. Comparison of results with other published results

Clinical trials		Years of life saved, YLS	ICER compared to no ART, US$/YLS
Reference	ART model	(unadjusted)	(unadjusted)
ACTG 320 Study (1997)	PI-based second-line	1.54	$21,000
Johns Hopkins HIV Clinic (1999)	PI-based second-line	1.63	$16,000
INCAS trial (1998)	NNRTI-based first-line	2.34	$12,000
Dupont 006 trial (1999)	NNRTI-based first-line	2.84	$13,000
Cohort simulations			
Individual-based	Starting criteria	YLS	ICER
Freedberg *et al.* (2001)			
	CD4 < 50 cells/mm3	1.45	$22,000
PI-based second-line	CD4 < 200 cells/mm3	1.99	$16,000
	CD4 < 500 cells/mm3	2.08	14,000
Goldie *et al.* (2006)			
One-line ART only	CD4 < 200, or CD4<350	3.96	$1100
Two lines of ART [i]	and 1 severe OD	4.25	$1200
Freedberg *et al.* (2007)			
	CD4 < 200 cells/mm3	2.33 [ii]	Not specified
NNRTI-based first-line only	CD4 < 250 cells/mm3	2.43 [ii]	$430
	CD4 < 350 cells/mm3	2.52 [ii]	$550
Two lines of ART: NNRTI-based first-line, then PI-based second line	CD4 < 200 cells/mm3	4.19 [ii]	Not specified
	CD4 < 250 cells/mm3	4.35 [ii]	$1,060
	CD4 < 350 cells/mm3	4.53 [ii]	$1,530

Average-based				
Dutta (2008, this study)		Till EOFU	Till TF	ICER: $/YLS [ii]
NNRTI-based first-line only	CD4 < 200 cells/mm3	2.62 [iii]	2.56 [iii]	$513 (till EOFU)
	CD4 < 250 cells/mm3	2.56 [iii]	2.51 [iii]	$494 (till EOFU)
Two lines of ART: NNRTI-based first-line, then PI-based second line	CD4 < 200 cells/mm3	4.31 [iii]		$1,025
	CD4 < 250 cells/mm3	4.26 [iii]		$979

ACTG, AIDS Clinical Trial Group; INCAS, Italy-Netherlands-Canada-Australia Study;
[i] First-line switching criteria: CD4 count drops 90%.   [ii] YLS from Freedberg *et al.* (2007), discounted at 3% p.a.
[iii] Quasi-YLS, undiscounted. Compared to Freedberg *et al.*, this is a calculation of average years of life gained across the cohort in the duration of treatment follow-up. Life years left for those alive at the end of follow-up could not be incorporated.

In a previous study where a two-line strategy was incrementally compared to a single-line strategy, it was found to be cost-effective by the WHO criteria of an ICER below three times the GDP per capita [2]. Table 15 provides an incremental cost-effectiveness comparison of single-line till EOFU against single-line till TF, and two-line ART against

single-line till TF. While single-line till EOFU is found to be cost-effective, the two-line strategy is not. This is a major policy-relevant departure from the published literature, even as this state-transition simulation model performs well in line with other studies on the rankings of treatment models and the value of ICER for all other comparisons (Table 14).

Table 15. ICER of treatment alternatives compared to ending single-line ART at TF

ART Model, starting criteria (N=122,947)	Years of follow-up	Premature deaths averted compared to one line ART till TF [i]	Total costs for cohort, till end of follow-up (US$ mil.)	Discounted mean YLS[ii]	Discounted average per-person total costs (US$)	**ICER** (US$/YLS), compared to one line ART till TF	Result
Cohort baseline CD4< 200 cells/mm^3							
One line of ART, NNRTI-based (till end of follow-up)	5	7,108	$164.3	2.33*	$1,450	$1,081	
Two lines of ART, NNRTI-based, then PI-based	7	12,250	$476.6	1.86	$6,412	$14,491[iv]	Dominated[iii]
Cohort baseline CD4< 250 cells/mm^3							
One line of ART, NNRTI-based (till end of follow-up)	5	6,892	$165.7	2.43*	$1,446	$1,108	Dominated
Two lines of ART, NNRTI-based, then PI-based	7	12,716	$476.9	1.92	$6,286	$13,545[iv]	Not cost-effective[v]

[i] Baselines are matched; for example, one line of ART till EOFU with CD4<200 is compared to one line of ART till TF, CD4<200.
[ii] *In the case of '1-line till EOFU', the YLS is that compared to no treatment, since the strategy has roughly the same overall life-expectancy as '1-line till TF'. In case of the 2-line strategy the YLS is the incremental discounted life-expectancy compared to the '1-line till TF' strategy. All YLS are based on life expectancies in Freedberg *et al.* (2007). Sensitivity tests using the YLS values in Table 14, calculated in this simulation but limited to the period of follow-up are provided in an appendix.
[iii] Weakly dominated; higher ICER than the next more costly strategy.
[iv] Freedberg *et al.* (2007) find two lines of ART cost-effective compared to one line; ICER of $1850 for both starting criteria.
[v] According to WHO guidelines, cost-effectiveness of health interventions in resource-poor settings involves an ICER less than three times the GDP per capita. Using the 2008 GDP per capita in India of $735 (Source: UNDP 2007), this limit is $2205.

Sensitivity analysis

Comprehensive sensitivity analysis of model inputs and assumptions was conducted, and some important results are presented here, with the rest in an appendix. The cost-effectiveness of a two-line strategy decreased with any increased tendency of physicians to switch ART for year 2 treatment failures, while remaining cost-effective.

Table 16. Sensitivity of results to stringency of physician-led definition of first-line TF

Proportion of first-line TF diagnosed in Yr 2, for CD4 strata >51 cells/mm^3	ICER (US$/YLS)		Deaths averted compared to no treatment	
	Single line of NNRTI-based ART, till TF, CD4<250	Two lines of ART, NNRTI-based till TF, then PI-based, CD4<250	Single line of NNRTI-based ART, till TF, CD4<250	Two lines of ART, NNRTI-based till TF, then PI-based, CD4<250
0.1	$440.7	$844	89,566	99,596
0.15	$439.8	$857	88,923	99,574
0.25	$438.1	$881	87,650	99,530
0.33	$436.6 (baseline)	$901	86,643 (baseline)	99,496
0.45	$434.4	$929	85,151	99,445
0.5	$433.5	$941	84,536	99,424
0.66	$430.6	$979 (baseline)	82,714	99,359 (baseline)
0.8	$427.7	$1,011	80,928	99,304
1	$423.7	$1,057	78,601	99,228

Notes: Yr, year. Baseline refers to the rate used for all results reported elsewhere, unless otherwise mentioned.
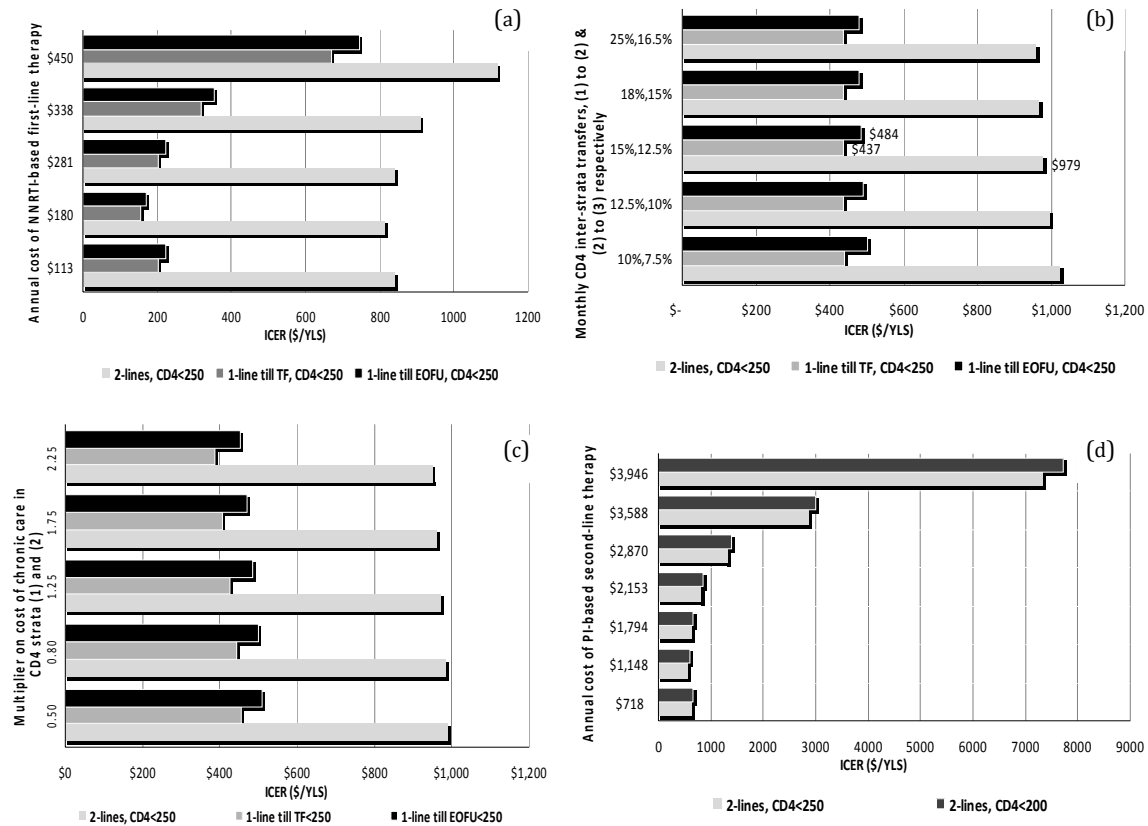
Fig. 9. Sensitivity of ICER to model inputs across select treatment models. (a) Variation in the annual cost of an NNRTI-based first-line treatment combination, from 50% to 200% of baseline =$225. (b) Variation in first-line ART efficacy. Baseline efficacy: inter-strata monthly transfer rate for movement from stratum (1) to (2) = 15%, & from (2) to (3) = 12.5%. Tested range of variation: from a reduction in both rates by ~35% to an increase by ~60%. (c) Variation in cost of chronic care per patient, from 50% to 225% of baseline cost for CD4 strata (1) and (2), as in Table 12. (d) Variation in the annual cost of a PI-based second-line treatment combination, from 50% to 275% of baseline =$1435.

Notes: 1-line till EOFU: First-line ART only, till end of follow-up. 1-line till TF: First-line ART only, stopped at treatment failure. 2-lines: First-line combination ART based on NNRTI, followed on failure by PI-based second-line regimen.

The price of PI-based second-line regimen had a significant impact on the cost-effectiveness of a two-line model, whereas increases in chronic costs of HIV disease slightly improved the incremental cost-effectiveness of ART against no treatment (working through the channel of moving people into higher CD4 strata where the costs are much lower). The results on ICER were not significantly sensitive to changes in the assumptions on the efficacy of first-line ART in causing inter-strata transfers related to improving immunological profile (Table 9). Separately, the sensitivities of results to cost of CD4 tests, and the efficacy of PI-based second-line regimen were also tested and not found to be large. The increasing rate of background immune decline due to viral resistance when physicians did not define or act on treatment failure (single-line till EOFU) led to its declining cost-effectiveness. At very high rates of such retrogression, 15% per CD4 stratum per month, the

single-line till EOFU strategy declined from 'very cost-effective' (ICER below GDP per capita of US$ 735) to 'cost effective'. Immune retrogression was also implemented in the PI-based second-line arm of a two-line strategy, but did not impact on its cost-effectiveness in a way that would change its rank. Values using the starting criteria variation of CD4 below 250 cells/mm³ are shown below (the dynamics were the same with the other starting criteria).

Table 17. Sensitivity to rate of immunological retrogression in certain treatment models

	ICER (US$/YLS)		Deaths averted compared to no treatment	
Immunological retrogression rate in absence of physician determined TF (see Table 11)	Single line of NNRTI-based ART, till EOFU, CD4<250	Two lines of ART, NNRTI-based till TF, then PI-based, CD4<250	Single line of NNRTI-based ART, till EOFU, CD4<250	Two lines of ART, NNRTI-based till TF, then PI-based, CD4<250
2.5%	$ 484.0	$977	96,048	99,874
4%	$ 494.0 (baseline)	$979 (baseline)	93,535 (baseline)	99,359 (baseline)
5%	$ 504.5	$981	90,971	98,756
7.5%	$ 538.6	$989	83,417	96,652
10%	$ 588.5	$1,002	74,146	93,597
15%	$ 744.4	$1,035	54,373	85,878

Discussion

The number of adult individuals eligible for antiretroviral treatment (ART) in India, in the beginning of 2008, is at least 1.87 million (77% of adult PLWHA). This is as per applying a plausible sectioning of the PLWHA population into CD4 strata, after accounting for the local age of the HIV epidemic, and applying an eligibility criterion. The eligibility requirement is interpreted conservatively, on only the immunological basis of the NACO (WHO) guidelines on ART eligibility, i.e., 'treatment required' for CD4 count less than 200 cells/mm³, and not including those with higher CD4 counts who would be eligible because of being in WHO clinical stage 4 or a qualifying stage 3 opportunistic disease condition (criteria which are important in resource-poor settings unable to afford extensive CD4 testing). Using the immunological criterion of 'recommended' rather than required treatment, i.e., CD4 count less than 250 cells/mm3, the number eligible falls to 1.73 million (71% of PLWHA). This value is numerically lower than other recent estimates [47] but higher on a percentage basis of total PLWHA. This represents two factors in play.

The first is a reduction in the estimated total number of PLWHA for 2008 to about 2.41 million, from a figure 2.37 times higher, or about 5.7 million, quoted in recent studies of ART policy [2]. The second is a reckoning captured in the skew in the PLWHA

proportions towards higher CD4 strata that the epidemic has matured in the states carrying the largest number of the PLWHA (called the 'old' states in this study). This factor is borne out in the plateau of the prevalence curves over time in Fig. 2 (a), as well as in recent studies [11], including those specifically looking at sero-stabilization in the major epidemic states of southern India [16]. Though there will be an incidence 'bump' in coming years from the 'new' epidemic states that calls for reinvigorating prevention efforts there, the overall policy imperative is for adequate and affordable treatment for the large cohort of HIV+ people, so as to best reduce the rates of debilitation and increase the life years available for those individuals to enjoy productively.

The free ART program which is currently being scaled up by the Government of India is a commendable effort to meet the requirements of those PLWHA seeking subsidized treatment and is also a source for support, treatment counseling and for firming up and maintaining adherence in those utilizing the prescribed therapy. However, using the best available estimates of those on such free, and in the case of NGOs, mostly or partially subsidized treatment, between 100,640-101,000 ART-eligible individuals are being served as of early 2008 (5.4% of total PLWHA eligible for treatment using the strict CD4 criteria). The rest of the 95% of eligible PLWHA either must purchase ART through out-of-pocket (OOP) funds – while also bearing the expense for diagnostics, chronic care inclusive of prophylactic drugs, and enhanced nutrition requirements – or go without treatment. Using an available large sample estimate of public and private sector rates of utilization of ART among patients presenting for generalized HIV-related care [19], this study established the number of OOP patients ART currently (approximately 719,000 individuals). The private sector ART utilization rate from the sample – which was derived from different sources across India - was assumed to apply for OOP patients across the nation. This assumption was sensitivity tested and found to impact the overall size of the cohort of untreated PLWHA.

The total annual spending in 2007-8 on first-line ARVs in free/subsidized and OOP treatment arms was estimated using the utilization rates above, after unit costs were established from both a current market survey in New Delhi (for private sector prices applicable to OOP patients) and as an average of prices applicable to bulk purchasers in the public sector based on published sources [2, 23, 25]. These estimates of spending suggest that public and NGO sector spending on ARVs is only 9% of OOP spending, and 8% of the

total market size, a low figure which should allow for substantial expansion in coverage if the expansion meets criteria of cost-effectiveness in a public health approach.

In the interest of evaluating the public health benefits of expanded coverage, the total deaths and severe opportunistic disease episodes were simulated in the cohort of individuals receiving no ARVs, though a portion of whom received prophylaxis using co-trimoxazole against opportunistic infection. The number of individuals in this national untreated cohort, using the numbers of PLWHA above and the total utilization rate of ART in free/subsidized and OOPS arms, was 1.592 million adults, without limitation on their CD4 status at the baseline date (assumed to be February 2008). Hence, the cohort included individuals at relatively earlier stages of HIV disease, i.e., CD4 count above 250 cells/mm^3, and those in whom disease had progressed further (CD4 count below 250 cells/mm^3). The proportion receiving opportunistic infection prophylaxis was 51%, as estimated using the associated utilization rates separately in public and private sector patients from the same source as for ART utilization [19]. Using the methods described in the text, and incorporating efficacy of co-trimoxazole prophylaxis in reducing monthly risk of opportunistic infection, the disease course in this cohort of 1.6 million individuals was simulated for a total of five years. At the end of the simulation period, the cohort in total had suffered 84.74% mortality, or about 1.35 million deaths.

Receiving only co-trimoxazole prophylaxis was instrumental in greatly reducing the number of severe opportunistic disease (OD) episodes and hence associated morbidity and mortality, but did not impact overall mortality rates which extended from generally declining immunological profile in the absence of ART (and background increase in viral load), and heightened chronic mortality rates in the cohort. However, an estimated 2,075 excess deaths would result if co-trimoxazole was withdrawn in the OI prophylaxis sub-cohort. These results follow other estimates of co-trimoxazole efficacy and findings of limited impact on life-expectancy but substantial benefits in terms of reducing the rate of severe infections which require additional hospitalization and care [2, 37]. The use of co-trimoxazole in a sub-cohort of 0.81 million in an overall cohort of 1.6 million individuals in India without ARVs was associated with a saving of approximately US$130 million in treatment costs for severe OD, at a cost of US$21 million, a more than six times return on investment. Considering the affordability and ease of use of co-trimoxazole (annual costs below US$10), there is a strong case for increasing the utilization of this option in India.

With the same simulation technique, but incorporating a model of ART efficacy in terms of improving the immunological profile of recruited treatment-naïve patients, the incremental cost-effectiveness of three ART models were estimated, allowing for two options in each related to the baseline CD4 count criteria for inclusion. The size of the cohort simulated was 122,947 adults, both male and female, or about 15% of the estimated HIV+ adult population in India receiving no ART or OI prophylaxis (i.e., no treatment at all). This number is also roughly in the range of the individuals likely to be receiving free/subsidized ART by mid-2008. Therefore, the hypothetical recruitment of such a cohort en masse at the beginning of February 2009, with their disease course simulated over a follow-up period of 5-7 years, represents an overall doubling of the scale of the free ART program of the Government of India which started in April 2004 during the National AIDS Control Program-II (NACP-II), in line with the necessary pace implied by the announcement under NACP-III (2007-2014) of achieving 340,000 patients – adults and children – receiving free ARVs by 2011 [48].

The incremental cost-effectiveness ratios (ICERs) presented in the text use the averted mortalities for the years of follow-up by treatment type from this simulation, valued at the discounted years of life saved (YLS) from the stochastic state-transition study by Freedberg *et al.* [2]. The present simulation analysis allowed the calculation of quasi-years of life saved (YLS), based on the mortality profile in the treated cohort only during the years of follow-up (5-7 years depending on the treatment model). The reasons for preferentially presenting the ICERs estimated using the YLS from Freedberg *et al.* were discussed in the methods section above. The ICERs calculated with the quasi-YLS are presented in the appendix for comparison. The quasi-YLS were close to those estimated by Freedberg *et al.*, once the latter discounted future life years, and using them to calculate the ICER does not change the cost-effectiveness rankings from those in Table 13-14, except between the two variations of a single-line ART model till treatment failure (the ICERs here become roughly equal). The incremental benefit under two lines of ART compared to one line of ART (with all ART stopped at first-line treatment failure) was an average of 1.8 quasi-YLS across the two baseline CD4 variations, compared to 1.9 YLS in Freedberg *et al.* [2].

As in previous studies for resource-poor settings [2, 5], the treatment model of a single line of ARVs – patterned on the NNRTI-based triple combination regimen with a nevirapine core widely used in India – was 'very' cost-effective (ICER less than GDP per capita) compared to no treatment. Within this treatment model, starting ART earlier in the

disease course in patients with CD4 T-cell counts below 250 cells/mm³ strongly dominated starting later with baseline CD4 below 200 cells/mm³. This is more support for starting earlier compared to Freedberg *et al.*, who found that this option weakly dominated starting later, i.e., the former had lower ICER but at a higher cost. However, the cost advantage in this simulation for starting earlier was small.

For the case of first-line treatment failure, studies have expressed an interest in modeling the decision to stop ART vs. continuing therapy till death or loss to follow-up. This is an important consideration for HIV policy in resource-poor countries where viral resistance and hence likely first-line failure in the course of therapy is an increasingly pervasive factor. The increase in resistance to first-line therapy for HIV-1 has been noted in India [44-46]. In resource-poor settings, an earlier tendency to first-line failure in treatment-naïve patients is occurring with the circulation of mutated strains in the population which have evolved after a long history of rote prescription of an older class of NRTI-based combination first-line mono- and duo-therapies, regardless of new patient clinical outcomes and presentation (whether due to poor availability of NNRTIs or alternative NRTIs/NtRTIs, or due to poor training of physicians). Separately to blame is patients' poor adherence while on the ARV regimens. The latter is suggested as the most important, though not unassailable, contributor to resistance emergence in India [46].

Treatment failure in some form for patients on ART – first-line and second-line – was considered in all the models in this simulation. In the absence of physician determined first-line treatment failure as per the strict observation of WHO guidelines, it was assumed that immunological deterioration would begin in a certain proportion of patients. The baseline assumption was of 4% retrogression in each CD4 stratum per patient month after a year on treatment, a parameter which was sensitivity tested and found to be significant. First-line therapy could be continued till end of follow-up (EOFU), in one type of treatment model, in which case all patients would continue to derive benefit from reduced incidence of opportunistic infection, and even those with progressive treatment failure would continue to be at lower risk for chronic mortality than without treatment. This treatment model would still be very cost-effective, with an ICER of US$-494-513 depending on variation in CD4 starting criteria. An ARV-conserving, public-health approach in the climate of limited availability of drugs would stop ART at the physician-determined point of first-line treatment failure, with the risk of such failure assumed in this study at various levels dependent on CD4 status. Though such a stop-not-switch treatment model would be

incrementally more cost-effective compared to first-line ART till EOFU, it would avert substantially fewer deaths, and public health officials should only consider it in situations of severe rationing of combination ARVs. When the additional benefits of lives saved under first-line ART till EOFU are considered, the ICER of this model compared to ending ART at first-line treatment failure was US$1,446-1450, still cost-effective by WHO criteria (less than three times per capita GDP).

Protease-inhibitor (PI) based second-line ART was made available as a treatment option for those with first-line treatment failure. The combination therapy for this arm of the cohort was patterned after a ritonavir-boosted PI such as indinavir, and two NRTIs, as commonly used in second-line treatment in India. Those without first-line treatment failure in this 'two line' model continued on the NNRTI-based combination. This two-line treatment model was also cost-effective compared to no treatment, and within it, the variation based on initially recruiting at an earlier stage of HIV disease (baseline CD4 counts below 250 cells/mm^3) had a lower ICER. Though two lines of ART increased the number of deaths averted – till seven years of follow-up – the increase was not sufficient to make it cost-effective compared to ending first-line treatment at EOFU, or at physician-determined treatment failure. This finding was robust to substantial reduction in the price of PI-based second-line combination regimens, and to various other sensitivity tests. This is a major departure from the findings of Freedberg *et al.* [2], though the ICER of a two-line compared to first-line only treatment – in the range of US$13,530-14,440 is still below the ICER of second-line treatment compared to no treatment in earlier studies (Table 14).

The results above were sensitive to several modeled assumptions, especially the cost of second-line ART. If costs of a PI-based second-line regimen within a two-line model would double, it would be difficult to construe such ART as cost-effective by WHO standards. Though an increase of this level is not a likely immediate occurrence in the present climate of negotiated ARV pricing for national treatment programs, the sensitivity to price is a matter of concern as new drug combinations – including those reducing the pill-burden by merging doses – come to physicians' notice in resource-poor countries. With resistance to standard combinations used in first-line therapy and as seen in sub-Saharan Africa, even to second-line PI-based therapy, attention will move to innovative yet expensive multi-class combinations that can be second-line or third-line choices (after failure of a PI-based second-line) and/or offer other benefits. New combinations without a PI, such as Bristol Myers-Squibb/Gilead's once-a-day *Atripla*™ (NRTI efavirenz, plus NtRTI

tenofavir and NRTI emtricitabine) reduce pill-burden and promise to increase adherence, but are introduced at a high cost in the West, and generics manufacturers may not offer much relief. Controversially, the price at which Cipla, a generics maker, is offering *Atripla*™ in India as *Viraday* has been assailed as too high by certain NGOs, citing an alleged price of US$1344 per patient year (40 rupees/US$)[49]. Cipla itself does not publish list prices.

The ICER for the two types of treatment models affected by physician decisions on treatment failure were differentially sensitive to the stringency of the definition of first-line failure for patients with CD4 counts above 50 cells/mm3 after the first year (i.e., when to stop ART or switch regimens based on immunological failure). While the incremental cost per YLS of stopping ART altogether did not change appreciably, the incremental costs of switching regimens increased with earlier switching without appreciable increase in lives saved. However, even with very early decisions to switch, the overall two-line policy remained cost-effective.

This study had several limitations, as discussed below. However, these should be considered against the fact that the tractability and customizability of this model executed in Microsoft Excel – compared to simulation models such as CEPAC [2, 5] coded in the programming language C++, which cannot be edited except by the developers – means that it can be used to evaluate policies at all levels of decision-making, and can be configured to suit the local PLWHA population's clinical profile, the appropriate efficacy parameters of the interventions, and local cost values. Qualitatively, the model also offers the advantage of specifically engaging with the issue of treatment failure and decisions to continue treatment or stop or to switch regimens, all valid policy considerations in resource-poor settings with a need to conserve ARV stocks while maintaining high adherence. Even without calibration to any other simulation platform, the model used here yields results within the prior established boundary on cost-effectiveness of ART from published sources. Over time, average-based state transition models such as this can be constantly improved, while offering a low cost/effort solution to policymakers who want a desktop tool to visualize the impact of various treatment strategies at different scales of expansion of coverage.

The specific opportunistic disease incidence risk, efficacy of co-trimoxazole as well as the CD4-independent effect of ART on mortality and incidence risk were not India-specific, and were sourced from various published studies and are subject to the limitations quoted therein [2, 5]. Certain such parameters from Freedberg *et al.* (who derive them from

a clinical trial in Côte d'Ivoire) stood out, especially the greater monthly chronic mortality risk in the lowest CD4 stratum (1), or below 50 cells/mm³, for those with no history of opportunistic disease. However, results were found only marginally sensitive to this and other parameters sourced externally. Additionally, the cost data other than on first-line ART pricing were based on the observational database also used by Freedberg *et al.*, and may have limited generalizability to an India-wide estimate. Toxicity and other treatment side-effects were not modeled may represent an adjustment to the quality of life under treatment, which would diminish the cost-effectiveness if quality-adjusted life years are used. Benefits of likely increase in total adherence rates in HIV+ individuals receiving ART from the aspects of increased counseling and structured care incorporated in the expansion of the free program were not estimated. There are factors with a combined, indistinct net effect on risk of transmission to HIV-negative partners from those receiving treatment: disinhibition (more risk) and reduced viral load (lower risk). These were estimated in a population-level model of India [47] but were outside the scope of this study. Addition of all these omitted effects in future iterations of this model would be worthwhile.

Certain classes of parameters were assumed in this study, especially on state transition of HIV disease and the efficacy of ART in delivering CD4 cell count increases. Though every effort was made to model the average rate of movements across CD4 strata – either in improving health or decreasing it as per context – so as to mimic the results from observational studies in India, the mechanism is quite distinct from that in stochastic, individual-based simulations that can implement patient transitions to calibrate to an overall rate of efficacy seen in clinical trials. However, the average-based results as in this study are defensible, in that they represent the dampening out of the individual-based variation – as would be seen in any stochastic simulation with a large enough N – towards population averages over time. Such a population-based technique is relevant when public health approach is used for HIV disease as suitable to a resource-poor setting rather than a specialized, individual-treatment focused model as in the West.

Several important treatment interventions in the face of increasing failure of first-line therapy remain to be effectively modeled for policymaker usage in the developing world, such as ritonavir-boosted PI (PI/r) in isolation as a possible alternative to a second-line combination regimen, structured treatment interruption, dose reduction, etc. Though these offer a significant challenge in modeling in a desktop software environment without recourse to a large number of assumptions (given the lack of efficacy studies from resource-

poor settings for some of these options other than PI/r), there is a great need for research to turn to these as the epidemics in India stabilize and patient histories mature.

An important finding of this study is that increasing survival comes at increasing cost. As India moves to generally adopt a two-line treatment model and the envelope of lives available to HIV+ patients expands, so will the requirements for the apparatus of chronic care, treatment counseling, and the total lifetime cost per-person borne outside the program (e.g., enhanced nutrition). The Government of India's latest request [50] of US$88 million to Round 7 of the Global Fund to Fight AIDS Tuberculosis and Malaria is focused on the need to increase the infrastructure of care, especially in rural areas, and to train healthcare workers and equip facilities to cater to the increasing demand. However, this investment is dwarfed by the treatment costs if a cohort the size of the proposed NACP-III target of 300,000 is offered a two-line model (CD4 below 250 cells/mm^3) with a realistic seven-year follow-up period. On the basis of this study such a strategy would cost nearly US$1.2 billion (undiscounted) in total. Though such a treatment model is programmatically cost-effective, at this scale the associated size of investment in India requires further analysis and debate given various other calls on public health resources.

References

1.   The Antiretroviral Therapy in Lower Income Countries Study Group. **Cohort Profile: Antiretroviral Therapy in Lower Income Countries (ART-LINC): international collaboration of treatment cohorts**. *Int. J. Epidemiol.* 2005,34:979-986.
2.   Freedberg KA, Kumarasamy N, Losina E, Cecelia AJ, Scott CA, Divi N, *et al.* **Clinical impact and cost-effectiveness of antiretroviral therapy in India: starting criteria and second-line therapy**. *AIDS* 2007,21:S117-S128.
3.   National AIDS Control Organization, Ministry of Health and Family Welfare, India Go. **Technical Report: India HIV Estimates-2006**. New Delhi; 2007.
4.   Zewdie D, De Cock K, Piot P. **Sustaining treatment costs: who will pay?** *AIDS* 2007,21:S1-4.
5.   Goldie SJ, Yazdanpanah Y, Losina E, Weinstein MC, Anglaret X, Walensky RP, *et al.* **Cost-Effectiveness of HIV Treatment in Resource-Poor Settings--The Case of Cote d'Ivoire**. *New England Journal of Medicine* 2006,355:1141.
6.   Freedberg KA, Losina E, Weinstein MC, Paltiel AD, Cohen CJ, Seage GR, *et al.* **The Cost Effectiveness of Combination Antiretroviral Therapy for HIV Disease**. *New England Journal of Medicine* 2001,344:824.
7.   Frieden TR, Das-Douglas M, Kellerman SE, Henning KJ. **Applying Public Health Principles to the HIV Epidemic**. *N Engl J Med* 2005,353:2397-2402.
8.   Grubb I, Perrièens J, Schwartleander B. **A public health approach to antiretroviral treatment: overcoming constraints**. Geneva: World Health Organization. Dept. of HIV/AIDS; 2003.
9.   Gilks CF, Crowley S, Ekpini R, Gove S, Perriens J, Souteyrand Y, *et al.* **The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings**. *Lancet* 2006,368:505-510.

10. National AIDS Control Organization, Ministry of Health and Family Welfare, India Go. **HIV/AIDS epidemiological Surveillance & Estimation Report for the Year 2005** 2006. Available from: http://www.nacoonline.org/fnlapil06rprt.pdf. Accessed: June 11 207.

11. Dandona L, Lakshmi V, Sudha T, Kumar GA, Dandona R. **A population-based study of human immunodeficiency virus in south India reveals major differences from sentinel surveillance-based estimates**. *BMC Medicine* 2006,4.

12. Dandona L, Lakshmi V, Kumar GA, Dandona R. **Is the HIV burden in India being overestimated?** *BMC Public Health* 2006,6:308.

13. National AIDS Control Organization, Ministry of Health and Family Welfare, India Go. **Annual HIV Sentinel Surveillance: Country Report 2006**. New Delhi: NACO; 2007.

14. International Institute for Population Sciences (IIPS) and Macro International. **National Family Health Survey (NFHS-3), 2005-06: India: Volume I**. Mumbai: IIPS; 2007.

15. Measure DHS (Macro International Inc.) and Indian Institute of Population Sciences. **India standard DHS 2005-06 (NFHS-3) dataset.** Measure DHS; 2007. Available from www.measuredhs.com.

16. Kumar R, Jha P, Arora P, Mony P, Bhatia P, Millson P, *et al.* **Trends in HIV-1 in young adults in south India from 2000 to 2004: a prevalence study**. *The Lancet* 2006,367:1164-1172.

17. Galárraga O, O'Brien ME, Gutiérrez JP, Renaud-Théry F, Nguimfack BD, Beusenberg M, *et al.* **Forecast of demand for antiretroviral drugs in low and middle-income countries: 2007-2008**. *AIDS* 2007,21:S97-S103.

18. Mahal A, Yazbeck AS, Peters DH, Ramana GNV. **The poor and health service use in India**. Washington, D.C.: The World Bank; 2001.

19. Vadrevu R, Rajasekaran S, Bollinger RC, Gupta A. **Knowledge, Attitudes, and Practices of Antiretroviral Therapy Among HIV-Infected Adults Attending Private and Public Clinics in India**. *AIDS Patient Care and STDs* 2007,21.

20. World Health Organization HIVAPGDG. **Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach. 2006 revision**.

21. Gupta I. **Willingness to Pay for Antiretroviral Therapy for HIV Positive Individuals in India**. *Forum for Health Economics & Policy* 2007,10:Economics of the HIV Epidemic, Art. 7.

22. Gupta I, Trivedi M, Kandamuthan S. **Costing of the Free ART Programme of the Government of India**. In.

23. Clinton Foundation HIV/AIDS Initiative (CHAI). **Antiretroviral (ARV) pricelist**. May 2007. http://www.clintonfoundation.org/pdf/chai-arv-price-list-050807.pdf Accessed: January 12 2008.

24. Clinton Foundation HIV/AIDS Initiative (CHAI), Pharmaceutical Services Team. **Engaging the ARV marketplace to optimize outcomes for patients**. Presentation. February 14, 2006. www.who.int/hiv/amds/(2.6)Clinton%20Foundation.ppt. Accessed: January 12 2008

25. Center for Social Markets, Business & Economic Development Research Project. **Report on the pharmaceutical sector in India**. 2004. Available from www.csmworld.org. Accessed: January 12 2007.

26. Indo-Asian News Service. **Patients to get second line of treatment** *The Tribune* December 1, 2007.

27. News agencies. **Tamil Nadu rolls out second-line AIDS treatment**; 2008. Available from: http://www.indiaenews.com/health/20080123/93440.htm. Accessed Feb 10, 2008.

28. Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, *et al.* **Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection**. *Ann Intern Med* 1997,126:946-954.

29. Kumarasamy N, Vallabhaneni S, Flanigan TP, Mayer KH, Solomon S. **Clinical profile of HIV in India**. *Indian Journal of Medical Research* 2005,121:377-394.

30. Fauci AS, Pantaleo G, Stanley S, Weissman D. **Immunopathogenic Mechanisms of HIV Infection**. *Annals of Internal Medicine* 1996,124:654.

31. Kumarasamy N, Venkatesh KK, Mayer KH, Freedberg K. **Financial burden of health services for people with HIV/AIDS in India**. *Indian J Med Res* 2007,126:509-517.

32. Rajasekaran S, Jeyaseelan L, Vijila S, Gomathi C, Raja K. **Predictors of failure of first-line antiretroviral therapy in HIV-infected adults: Indian experience**. *AIDS* 2007,21:S47-S53.

33. Cecelia AJ, Christybai P, Anand S, Jayakumar K, Gurunathan T, Vidya P, *et al.* **Usefulness of an observational database to assess antiretroviral treatment trends in India**. *National Medical Journal of India* 2006,19:14.
34. Ghate MV, Tripathy SP, Kumar BK, Godbole SV, Chittake A, Nyayanirgune P, *et al.* **Rate of hospitalization and inpatient care costs for HIV-1-infected patients in Pune, India**. *National Medical Journal of India* 2006,19:10.
35. Kumarasamy N, Solomon S, Chaguturu SK, Cecelia AJ, Vallabhaneni S, Flanigan TP, Mayer KH. **The changing natural history of HIV disease: before and after the introduction of generic antiretroviral therapy in southern India**. *Clin Infect Dis* 2005,41:1525-1528.
36. Kumarasamy N, Solomon S, Flanigan TP, Hemalatha R, Thyagarajan SP, Mayer KH. **Natural history of human immunodeficiency virus disease in southern India**. *Clin Infect Dis* 2003,36:79-85.
37. Yazdanpanah Y, Losina E, Anglaret X, Goldie SJ, Walensky RP, Weinstein MC, *et al.* **Clinical impact and cost-effectiveness of co-trimoxazole prophylaxis in patients with HIV/AIDS in Côte d'Ivoire: a trial-based analysis**. *AIDS* 2005,19:1299-1308.
38. Over M, Heywood P, Gold J, Gupta I, Hira S, Marseille E. **HIV/AIDS Treatment and Prevention in India: Modeling the Cost and Consequences**. Washington DC: The World Bank; 2004.
39. Krishnan A, Solomon S, Mayer KH. **Barriers and Facilitators to Antiretroviral Medication Adherence Among Patients with HIV in Chennai, India: A Qualitative Study**. *AIDS Patient Care and STDs* 2005,19.
40. Maru D, Khakha D, Tahir M, Basu S, Sharma S. **Poor follow-up rates at a self-pay northern Indian tertiary AIDS clinic**. *International Journal for Equity in Health* 2007,6:14.
41. Kumarasamy N, Vallabhaneni S, Cecelia AJ, Yepthomi T, Balakrishnan PPD, Saghayam SMS, *et al.* **Reasons for modification of generic highly active antiretroviral therapeutic regimens among patients in southern India**. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2006,41:53-58.
42. National AIDS Control Organization, Ministry of Health and Family Welfare, India Go. **Antiretroviral therapy guidelines for HIV-infected adults and adolescents including post-exposure prophylaxis**: NACO; 2007.
43. Losina E, Yazdanpanah Y, Deuffic-Burban S, Wang B, Wolf LL, Messou E, *et al.* **The independent effect of highly active antiretroviral therapy on severe opportunistic disease incidence and mortality in HIV-infected adults in Côte d'Ivoire**. *Antiviral Therapy* 2007,12:543-551.
44. Hira SK, Panchal K, Parmar PA, Bhatia VP. **High resistance to antiretroviral drugs: the Indian experience**. *International Journal of STD & AIDS* 2004,15:173-177.
45. Sachdeva N, Sehgal S, Arora SK. **Frequency of Drug-Resistant Variants of HIV-1 Coexistent With Wild-Type in Treatment-Naive Patients of India**. *MedGenMed* 2005,7:68.
46. Balakrishnan P, Shanmugam S, Nagalingeswaran K, Solomon S. **HIV-1 drug resistance among untreated patients in India: Current status**. *Journal of Postgraduate Medicine* 2006,52:183-186.
47. Over M, Marseille E, Sudhakar K, Gold J, Gupta I, Indrayan A, *et al.* **Antiretroviral therapy and HIV prevention in India: modeling costs and consequences of policy options**. *Sex Transm Dis* 2006,33:S145-S152.
48. The World Bank. **Project appraisal document on a proposed credit in the amount of SDR 167.9 million to the Republic of India for a Third National HIV/AIDS Control Project**. Washington, D.C.: The World Bank; 2007.
49. AIDS Healthcare Foundation. **Profit at what cost? AIDS drugs for all**; 2007. Available from: www.aidshealth.org. Accessed December 5 2007.
50. Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM). **Round 7 proposal of the Country Coordinating Mechanism, India, Preliminary.**: GFATM; 2007.

Paper III: A Comprehensive Look at Antiretroviral Therapy in India: Demand, Utilization, Impact of Non-Availability, and the Cost-Effectiveness of Different Treatment Options Including Second-line Therapy
Supporting Materials (Appendix)

Table A. Background data to Table 1 (main text): prevalence re-estimation for 2008

State	Type of epidemic (NACO-06 data)	RCHHLS 02-04: SLI of pub. sector users	Adj. bracket	Sent. Surv. Factor	NFHS3 factor: NFHS-3% ÷ANC06%	Sent. Surv. factor weight	NFHS factor weight	Adj. factor
Andhra Pradesh	Rural-Urban	Low	II	0.95	0.72	1	2	0.80
Karnataka	Rural-Urban	Low	II	0.95	0.68	1	2	0.77
Maharashtra	Rural-Urban	Low	II	0.95	0.78	1	2	0.83
Manipur	Rural-Urban	High	VIII	1.05	0.95	1	2	0.98
Nagaland	Rural-Urban	Med	V	1.00	N/A	1	2	1.00
Tamil Nadu	Rural-Urban	Low	II	0.95	0.76	1	2	0.82
Uttar Pradesh	Rural-Urban	Med	V	1.00	0.37	1	2	0.58
A & N Islands	Urban	Med	IV	0.95	N/A	2	1	0.95
Arunachal Pradesh	Rural	High	IX	1.15	0.00	2	1	0.77
Assam	Rural	Med	VI	1.05	6.29	2	1	2.80
Bihar	Rural-Urban	Med	V	1.00	0.00	2	1	0.67
Chandigarh	Urban	Low	I	0.85	N/A	2	1	0.85
Chhattisgarh	Rural-Urban	Med	V	1.00	0.27	2	1	0.76
Delhi	Urban	Low	I	0.85	0.60	2	1	0.77
Goa	Rural-Urban	Low	II	0.95	0.48	2	1	0.79
Gujarat	Rural-Urban	Low	II	0.95	0.29	2	1	0.73
Haryana	Rural	Med	VI	1.05	0.73	2	1	0.94
Himachal Pradesh	Rural	Med	VI	1.05	6.10	2	1	2.73
Jammu & Kashmir	Urban	Med	IV	0.95	6.86	2	1	2.92
Jharkhand	Rural-Urban	Med	V	1.00	0.60	2	1	0.87
Kerala	Rural	Low	III	1.00	0.00	2	1	0.67
Madhya Pradesh	Rural-Urban	Med	V	1.00	2.16	2	1	1.39
Meghalaya	Urban	High	VII	1.00	0.00	2	1	0.67
Mizoram	Rural-Urban	High	VIII	1.05	0.95	2	1	1.02
Orissa	Rural-Urban	Med	V	1.00	0.48	2	1	0.83
Pondicherry	Rural-Urban	Med	V	1.00	N/A	2	1	1.00
Punjab	Rural-Urban	Low	II	0.95	2.48	2	1	1.46
Rajasthan	Rural-Urban	Med	V	1.00	0.25	2	1	0.75
Sikkim	Rural	High	IX	1.15	0.00	2	1	0.77
Tripura	Urban	High	VII	1.00	0.00	2	1	0.67
Uttaranchal	Urban	Med	IV	0.95	0.00	2	1	0.63
West Bengal	Urban	Low	I	0.85	0.40	2	1	0.70

Notes: 'Sent. Surveillance factor': sentinel surveillance factor. This is based on the adjustment bracket provided to its left.

Table B. Averted deaths and cost-effectiveness, using quasi-YLS estimated in the simulation

ART models and starting criteria (N=122,947)	Years of follow-up[i]	Premature deaths averted compared to 'no ART'	Total costs for cohort, till end of follow-up (US$ mil.)	Un-discounted quasi-YLS compared to no treatment[ii]	Discounted average per-person total costs till EOFU (2008 US$)	ICER (US$/YLS), compared to no treatment	Result[v]
No ART	5	-	$52.9-53.2[iii]	2.875	$1,248	-	-
No ART, co-trimoxazole	5	127-169[iii]	$29.9-31.3[iii]	2.9	$959	N/C	-
One line of ART, NNRTI-based (till end of follow-up)							
Baseline CD4< 200 cells/mm³	5	93,401	$164.3	2.62	$1,450	$455	Dominated
Baseline CD4< 250 cells/mm³	5	93,535	$165.7	2.56	$1,446	$469	Dominated
One line of ART, NNRTI-based (till treatment failure)							
Baseline CD4< 200 cells/mm³	5	86,294	$146.4	2.9	$1,305	$414	
Baseline CD4< 250 cells/mm³	5	86,643	$147.1	2.9	$1,296	$415	Dominated
Two lines of ART, NNRTI-based, then PI-based[iv]							
Baseline CD4< 200 cells/mm³	7	98,543	$476.6	4.31	$6,412	$996	Dominated
Baseline CD4< 250 cells/mm³	7	99,359	$476.9	4.35	$6,286	$987	

EOFU, End of Follow-Up period (5-7 years, depending on treatment model); TF, Treatment Failure; mil., millions; YLS, Years of Life Saved; ICER, Incremental Cost-Effectiveness Ratio = (total cost till EOFU)/YLS; N/C: Not calculated

[i] For all treatment or no treatment variations, during the years of follow-up, bi-annual CD4 tests, chronic care management, terminal care, and severe opportunistic infection care are offered to all patients and costs incurred.
[ii] YLS based on mortalities over time in this simulation.
[iii] Depending on baseline CD4 characteristics of recruited cohort (higher values are for baseline CD4<250 cells/mm3)
[iv] Compared to no treatment, no OI prophylaxis, with 7 years of follow-up. This ensures that extending life under a two-line strategy is not penalized in terms of ICER by comparing with a shorter follow-up in the no treatment variation.
[v] All of these were strongly dominated on the basis of average per person costs till EOFU: higher ICER and higher cost.

Table C. ICER of alternatives compared to ending single-line ART at TF, using quasi-YLS

ART Model, starting criteria (N=122,947)	Years of follow-up	Premature deaths averted compared to one line ART till TF[i]	Total costs for cohort, till end of follow-up (US$ mil.)	Un-discounted mean quasi-YLS[ii]	Discounted average per-person total costs (US$)	ICER (US$/YLS), compared to one line ART till TF	Result
Cohort baseline CD4< 200 cells/mm³							
One line of ART, NNRTI-based (till end of follow-up)	5	7,108	$164.3	2.62*	$1,450	$960	
Two lines of ART, NNRTI-based, then PI-based	7	12,250	$476.6	1.78	$6,412	$15,142[iv]	Dominated[iii]
Cohort baseline CD4< 250 cells/mm³							
One line of ART, NNRTI-based (till end of follow-up)	5	6,892	$165.7	2.56*	$1,446	$1,052	Dominated
Two lines of ART, NNRTI-based, then PI-based	7	12,716	$476.9	1.79	$6,286	$14,491[iv]	Not cost-effective[v]

[i] Baselines are matched; for example, one line of ART till EOFU with CD4<200 is compared to one line of ART till TF, CD4<200.
[ii] Quasi-YLS based on mortalities over time in this simulation. *In the case of '1-line till EOFU', the YLS used is that compared to no treatment, since the strategy has roughly the same overall undiscounted life-expectancy as '1-line till TF'. In case of the 2-line strategy, the YLS is the incremental undiscounted life-expectancy compared to the '1-line till TF' strategy.
[iii] Weakly dominated; higher ICER than the next more costly strategy.
[iv] Freedberg *et al.* (2007) find two lines of ART cost-effective compared to one line; ICER of $1850 for both starting criteria.
[v] According to WHO guidelines, cost-effectiveness of health interventions in resource-poor settings involves an ICER less than three times the GDP per capita. Using the 2008 GDP per capita in India of $735 (Source: UNDP 2007), this limit is $2205.

Variable	SINGLE LINE TILL EOFU, BASE CASE RESULTS					
	NNRTI+ cotrix. OI ART < 200 cells/muL	NNRTI + cotrix. OI ART < 250 cells/muL	OI Prophylaxis only ART < 200 cells/muL	OI Prophylaxis only ART < 250 cells/muL	No Treatment ART < 200 cells/muL	No Treatment ART < 250 cells/muL
Deaths	**26,399 25,01**	**9**	**119,673 118,3**	**85**	**119,800**	**118,554**
60 months mortality%	21%	20%	97.34%	96.29%	97.44%	96.43%
Number of lives saved	**93,401 93,53**	**5**	**127**	**169**	N/A	N/A
without independent effect	72,382	72,730	N/A	N/A	N/A	N/A
No. of severe OI episodes	66,527	64,993	102,693	103,850	274,353	277,112
Total Cost (US$)	**164,285,133**	**165,652,902**	**29,898,109 31,26**	**3,434**	**52,916,154**	**53,221,501**
Total incremental cost (comp. to No Treatment)	111,368,978	112,431,401	(23,018,045)	(21,958,067)	N/A	N/A
Year 1 costs	38,039,752	38,012,319	13,244,790	13,107,546	23,580,548	22,699,827
Year 2 costs	33,424,323	33,697,785	8,721,802	8,997,941	15,462,723	15,396,748
Year 3 costs	31,955,478	32,299,105	4,713,106	5,194,028	8,279,269	8,694,161
Year 4 costs	30,853,035	31,233,482	2,240,035	2,679,209	3,901,553	4,379,299
Year 5 costs	30,012,544	30,410,212	978,376	1,284,710	1,692,062	2,051,466
Cost of ART	116,548,187	117,860,997	N/A	N/A	N/A	N/A
Cost of co-trimoxazole.	4,920,923	4,976,353	1,962,175	2,084,857	N/A	N/A
Cost of CD4 tests	7,975,647	7,791,762	4,946,270	5,537,552	30,143,446	29,811,313
Costs chronic care	271,641	257,445	1,231,440	1,218,178	1,232,742	1,219,917
Costs of OI severe episode treatments	25,025,958	25,123,738	17,528,780	17,964,702	17,357,555	17,781,776
Costs of terminal care	9,542,776	9,642,607	4,229,443	4,458,144	4,182,412	4,408,495
ICER (US$/YLS)	**513**	**494**				

Variable	SINGLE LINE TILL TF, BASE CASE RESULTS					
	NNRTI+co-trix. OI		OI Prophylaxis only		No Treatment	
	ART < 200 cells/muL	ART < 250 cells/muL	ART < 200 cells/muL	ART < 250 cells/muL	ART < 200 cells/muL	ART < 250 cells/muL
No. of individuals	122,947	122,947	122,947	122,947	122,947	122,947
Deaths 33,50	**6**	**31,911**	**119,673 118,3**	**85 119,8**	**00**	**118,554**
in ART +co-trix. OI	19,571	18,502	N/A	N/A	N/A	N/A
in 1st Line treat. Failure (no ART)	13,935	13,409	N/A	N/A	N/A	N/A
60 months mortality rate	27%	26%	97.34%	96.29%	97.44%	96.43%
Number of lives saved	**86,294 86,64**	**3**	**127**	**169**	**N/A**	**N/A**
without independent effect	70,282	70,773	N/A	N/A	N/A	N/A
No. of severe OI episodes	65,825	64,203	102,693	103,850	274,353	277,112
Total Cost (US$)	**146,415,471**	**147,074,391**	**29,898,109 31,26**	**3,434**	**52,916,154**	**53,221,501**
Total incremental cost (comp. to No Treatment)	93,499,317	93,852,890	(23,018,045)	(21,958,067)	N/A	N/A
Year 1 costs	38,055,966	38,015,709	13,244,790	13,107,546	23,580,548	22,699,827
Year 2 costs	32,073,559	32,318,724	8,721,802	8,997,941	15,462,723	15,396,748
Year 3 costs	28,367,436	28,566,728	4,713,106	5,194,028	8,279,269	8,694,161
Year 4 costs	25,200,410	25,341,649	2,240,035	2,679,209	3,901,553	4,379,299
Year 5 costs	22,718,100	22,831,582	978,376	1,284,710	1,692,062	2,051,466
Cost of ART	98,848,549	100,211,459	N/A	N/A	N/A	N/A
Cost of co-trimoxazole.	4,859,907	4,231,151	1,962,175	2,084,857	N/A	N/A
Cost of CD4 tests	8,170,953	7,928,428	4,946,270	5,537,552	30,143,446	29,811,313
Costs chronic care	344,781	328,363	1,231,440	1,218,178	1,232,742	1,219,917
Costs of OI severe episodes treatments	24,739,878	24,820,849	17,528,780	17,964,702	17,357,555	17,781,776
Costs of terminal care	9,451,402	9,554,142	4,229,443	4,458,144	4,182,412	4,408,495
ICER (US$/YLS)	**456 437**					

Variable	TWO LINES, BASE CASE RESULTS					
	NNRTI+PI+OI	NNRTI+PI+OI	OI Prophylaxis only	OI Prophylaxis only	No Treatment	No Treatment
	ART < 200 cells/muL	ART < 250 cells/muL	ART < 200 cells/muL	ART < 250 cells/muL	ART < 200 cells/muL	ART < 250 cells/muL
No. 1st line treatment failures	55,364	55,243	N/A	N/A	N/A	N/A
Deaths	**23,936 22,785**		**121,911**	**121,589**	**122,479**	**122,144**
in 1-st line ART + co-trix. OI	20,294	19,256	N/A	N/A	N/A	N/A
In 2nd Line + co.trix OI	3,642	3,529	N/A	N/A	N/A	N/A
60 months mortality rate	19%	19%	99.16%	98.90%	99.62%	99.35%
Number of lives saved	**98,543 99,359**		**22**	**36**	**N/A**	**N/A**
Without independent effect	73,855	75,989	N/A	N/A	N/A	N/A
No. of severe OI episodes	68,538	67,015	104,625	105,164	279,213	280,742
Total Cost	**476,586,907**	**476,922,446**	**30,458,155**	**32,106,173**	**53,196,748**	**53,822,426**
Total incremental cost	423,390,159	423,100,020	N/A	N/A	N/A	N/A
Year 1 costs	38,051,933	38,023,968	13,242,549	13,104,870	23,578,380	22,697,231
Year 2 costs	46,473,789	46,444,405	8,721,802	8,997,941	14,779,753	14,688,327
Year 3 costs	61,385,771	61,286,460	4,713,106	5,194,028	8,279,269	8,694,161
Year 4 costs	72,614,394	72,576,822	2,240,035	2,679,209	3,901,553	4,379,299
Year 5 costs	80,502,088	80,567,662	978,376	1,284,710	1,692,062	2,051,466
Year 6 costs	86,438,916	86,615,854	403,084	586,562	693,169	916,039
Year 7 costs	91,120,017	91,407,275	159,202	258,853	272,563	395,903
Cost of 1st Line ART	117,447,371	119,431,920	N/A	N/A	N/A	N/A
Cost of 2nd Line ART	297,234,122	295,421,698	N/A	N/A	N/A	N/A
Cost of co-trimoxazole.	6,926,641	6,998,434	1,991,410	2,128,421	N/A	N/A
Cost of CD4 tests (bi-annual)	8,252,466	8,069,137	5,042,053	5,727,542	30,687,200	30,532,657
Costs of chronic care	246,297	234,457	1,260,082	1,256,488	1,260,308	1,256,862
Costs of OI treatment	33,103,720	33,258,470	17,867,679	18,436,889	17,685,031	18,237,973
Costs of terminal care	13,376,289	13,508,328	4,296,931	4,556,833	3,564,208	3,794,934
ICER (US$/YLS)	**1,025 979**					